

H II regions in the CALIFA survey: I. catalogue presentation

C. Espinosa-Ponce¹,¹★ S. F. Sánchez,¹ C. Morisset,² J. K. Barrera-Ballesteros,¹
L. Galbany³,³ R. García-Benito⁴,⁴ E. A. D. Lacerda⁵¹ and D. Mast^{5,6}

¹*Instituto de Astronomía, Universidad Nacional Autónoma de México, AP 70264, 04510 Mexico City, Mexico*

²*Instituto de Astronomía, Universidad Nacional Autónoma de México, AP 106, 22800 Ensenada, B. C., Mexico*

³*Departamento de Física Teórica y del Cosmos, Universidad de Granada, E-18071 Granada, Spain*

⁴*Instituto de Astrofísica de Andalucía, CSIC, Apartado de correos 3004, E-18080 Granada, Spain*

⁵*Universidad Nacional Córdoba, Observatorio Astronómico de Córdoba, Laprida 854, (X5000BGO) Córdoba, Argentina*

⁶*Consejo de Investigaciones Científicas y Técnicas de la República Argentina, Avda. Rivadavia 1917, C1033AAJ CABA, Argentina*

Accepted 2020 March 12. Received 2020 March 12; in original form 2019 November 20

ABSTRACT

We present a new catalogue of H II regions based on the integral field spectroscopy (IFS) data of the extended CALIFA and PISCO samples. The selection of H II regions was based on two assumptions: a clumpy structure with high contrast of H α emission and an underlying stellar population comprising young stars. The catalogue provides the spectroscopic information of 26 408 individual regions corresponding to 924 galaxies, including the flux intensities and equivalent widths of 51 emission lines covering the wavelength range between 3745 and 7200 Å. To our knowledge, this is the largest catalogue of spectroscopic properties of H II regions. We explore a new approach to decontaminate the emission lines from diffuse ionized gas contribution. This diffuse gas correction was estimated to correct every emission line within the considered spectral range. With the catalogue of H II regions corrected, new demarcation lines are proposed for the classical diagnostic diagrams. Finally, we study the properties of the underlying stellar populations of the H II regions. It was found that there is a direct relationship between the ionization conditions on the nebulae and the properties of stellar populations besides the physical condition on the ionized regions.

Key words: techniques: spectroscopic – ISM: general – H II regions – galaxies: ISM – galaxies: star formation.

1 INTRODUCTION

Classical H II regions are clouds of ionized gas surrounding young massive stars in which the star formation has recently taken place (<15 Myr). The ultraviolet photons that ionize the surrounding medium are produced in large amounts by OB stars that were formed in the H II regions. Since OB stars only live ~ 10 Myr, the H II regions trace recent star formation processes. These regions possess a broad range of physical size (diameter), from a few parsecs, such as the Orion nebula with $D \lesssim 8$ pc (Anderson 2014), to hundreds of parsecs, like 30 Doradus with $D \sim 20$ pc, NGC 604 with $D \sim 460$ pc, or NGC 5471 with $D \sim 1$ kpc (Oey et al. 2003; García-Benito et al. 2011). The smaller regions are ionized by a single star or a handful of them, while the largest ones are ionized by a young cluster of thousands of them. Those larger star-forming (SF) complexes are the precursors of the extragalactic giant H II regions found commonly in the discs of spiral galaxies (e.g. Hodge & Kennicutt 1983; Dottori 1987; Dottori & Copetti 1989; Knapen

1998), or in starbursts and blue compact galaxies (Kehrig et al. 2008; López-Sánchez & Esteban 2009; Cairós et al. 2012).

In addition to H II regions, there are other emission-line objects with different main excitation mechanisms: planetary nebulae and objects photoionized by hard radiation fields such as active galactic nucleus (AGN), shock excitation, or even post-AGB stars (e.g. Veilleux & Osterbrock 1987; Binette et al. 1994; Stasińska & Izotov 2003; Binette et al. 2009; Morisset & Georgiev 2009; Flores-Fajardo et al. 2011; Papaderos et al. 2013; Singh et al. 2013). All these objects, including H II regions, could be classified, in principle, based on their emission line ratios. The first diagnostic diagram using the emission line ratios [N II] λ 6583/H α versus [O III] λ 5007/H β was proposed by Baldwin, Phillips & Terlevich (1981, hereafter BPT diagram). Osterbrock (1989) and Veilleux & Osterbrock (1987) proposed a demarcation curve between AGN and SF galaxies and two new diagnostic diagrams.¹ Theoretical

¹The two more commonly used diagnostic diagrams are [S II] λ 6717, 6761/H α versus [O III] λ 5007/H β and [O I] λ 6300/H α versus [O III] λ 5007/H β .

* E-mail: cespinoza@astro.unam.mx

photoionization models have been used to infer the demarcation line between the different ionization sources.

Dopita et al. (2000) and Kewley et al. (2001) proposed one of the first classification schemes to identify pure galaxies hosting star formation and AGNs, using a combination of stellar population synthesis and photoionization models. Their models assume that the emission-line spectrum depends directly on: (i) the chemical abundances of heavy elements (oxygen being the most important) of the gas phase within an H II region; (ii) the shape of ionizing radiation spectrum; (iii) the electron density; and (iv) the geometrical distribution of gas with respect to the ionizing sources. All geometrical factors are included in the so-called ionization parameter q (defined as the ratio of ionizing photon density to hydrogen density) or its dimensionless counterpart, U .² They assume a priori that these parameters are independent one from each other and, hence, the models are presented as grids of the four of them. On the other hand, Kauffmann et al. (2003) use the Sloan Digital Sky Survey (SDSS; York et al. 2000) data to propose a different demarcation line to separate ionization by SF and AGNs, based purely on observational properties.

H II regions allow the present chemical composition of the interstellar medium (ISM) to be determined. They have been used for a long time to measure the gas-phase abundance directly at discrete locations in galaxies (e.g. Diaz 1989; Rosales-Ortega et al. 2011; Marino et al. 2013; Sánchez-Menguiano et al. 2016). These measurements provide the necessary information to understand the distribution of chemical abundances across the optical extension of galaxies, to constrain different theories of galactic chemical evolution, to obtain information about the stellar nucleosynthesis in SF galaxies, and to derive the star formation histories of galaxies.

Classical studies of the chemical abundance of extragalactic nebulae are based on the observations of nebular emissions through long-slit or single-aperture spectrographs. Unfortunately, these techniques provided a limited sample of galaxies in detail, including a low number of H II regions per galaxy, in general. Therefore, there was a limited coverage of these regions along the whole galaxy extension (e.g. Bresolin 2019). Conversely, the classical slit-spectroscopy surveys introduce a bias by aperture effects: long-slit observations do not integrate over the full extension of an H II region. Instead, only the brightest region of each nebula is observed, i.e. the areas of higher excitation. Hence, the derived spectra may not be representative of the whole ionization conditions (e.g. Dopita et al. 2014).

Moreover, traditional surveys could present a biased sample of H II regions. Most of these spectroscopic surveys are biased towards high-contrast (thus, equivalent width of emission lines) nebulae located in the outer disc of galaxies. Since late-type galaxies offer the best contrast with their lower surface brightness in the continuum, these types of galaxies were mainly observed by these surveys. Thus, the final samples and conclusions could be biased to the properties of H II regions in a particular type of galaxy. Moreover, in most of the cases, the studies were restricted to regions ionized by younger and/or massive stellar clusters. In general, the central regions of galaxies were excluded, although it is known that they distinguish themselves spectroscopically from those in the disc: inner SF regions present stronger low-ionization forbidden emission

and may populate the right branch of the BPT diagram (Kennicutt, Keel & Blaha 1989; Ho, Filippenko & Sargent 1997; Sánchez et al. 2012b).

Despite these caveats, these studies provide important relations, clear patterns, and scaling laws, e.g. characteristics versus integrated abundances: Moustakas & Kennicutt 2006; effective yield versus luminosity and circular velocity relations (e.g. Garnett 2002); abundance gradients and the effective radius of discs (e.g. Diaz 1989); possible differences in the gas-phase abundance gradients between normal and barred spirals (Martin & Roy 1994; Zaritsky, Kennicutt & Huchra 1994); surface brightness versus metallicity relations, luminosity–metallicity, and mass–metallicity (Lequeux et al. 1979; Skillman, Kennicutt & Hodge 1989; Vila-Costas & Edmunds 1992; Zaritsky et al. 1994; Tremonti et al. 2004), to cite just a few.

In the last years, new observational techniques, such as integral field spectroscopy (IFS) with a large field of view (FoV) and multiobject spectroscopy, allow us to observe hundreds of H II regions per galaxy and a full two-dimensional (2D) coverage of the disc of nearby ones (e.g. Rosales-Ortega et al. 2010; Sánchez et al. 2012a, 2014). This new generation of emission-line surveys provided large catalogues with thousands of objects over an unbiased sample of galaxies, from early- to late-type galaxies (Marino et al. 2016; Sánchez-Menguiano et al. 2016). Moreover, the sampled H II regions are at any galactocentric distances (even at more than $2 R_e$),³ which allows us to study a wider range of H II regions.

The Calar Alto Legacy Integral Field Area survey (CALIFA; Sánchez et al. 2012a) was one of the pioneering in the massive exploration of H II regions over complete samples of galaxies (e.g. Sánchez et al. 2014). CALIFA provides IFS data for a well-defined and statistically significant sample of galaxies on the nearby Universe ($z < 0.08$). The observations cover the full optical extent (up to $3-4R_e$) of the galaxies of any morphological type and with a spectral range 3745–7200 Å. CALIFA still provides the best physical resolution (~ 0.8 kpc; Sanchez 2019) compared to other IFS galaxy surveys (e.g. Croom et al. 2012; Bundy et al. 2015) due to its limited redshift range and lower average redshift. However, it is overpassed by IFS observations with MUSE (e.g. López-Coba et al. submitted); none the less, this instrument does not cover wavelengths bluer than 4650 Å. In Sánchez et al. (2012a), the first H II region catalogue derived from the observations of the pilot study of the CALIFA survey was presented (p-CALIFA; Mármol-Queraltó et al. 2011). Besides that, several studies were performed with a larger statistical sample of H II regions based on later catalogue updates: oxygen abundance profiles on face-on spiral galaxies (Marino et al. 2016; Sánchez-Menguiano et al. 2016), global and local $M-Z$ relations and the dependence on the star formation rate (SFR) and the specific star formation rate (sSFR; Sánchez et al. 2013, 2017), etc. However, there was no public update of the CALIFA H II region catalogue after Sánchez et al. (2012a).

In this study, we present the updated CALIFA catalogue of H II regions for ~ 1000 nearby galaxies observed so far within the framework of this survey (2019 August). We propose new criteria to select H II regions within each galaxy and a new methodology to extract their main spectroscopic information. We explore the distribution of H II regions along the classic diagnostic diagrams. For the first time we do not use those diagrams to classify H II regions. Instead, we present an unbiased exploration of the location

²The ionization parameter is defined as $U = S/(4\pi r^2 n_H c)$, where S is the number of ionization photons emitted by a source per unit of time, r is the distance between the source and the nebula, n_H is the H density, and c is the speed of light.

³The effective radius, R_e , of a galaxy is defined as the radius at which half of the total light of the galaxy is emitted.

of these objects across the diagnostic diagrams. Moreover, we propose a new approach to correct the contamination by diffuse ionized gas (DIG). With the final catalogue of H II regions corrected by diffuse gas, we explore how this correction changes the distribution along those diagnostic diagrams. Finally, we show how H II regions distribute among galaxies of different morphologies and at different galactocentric distances, revising patterns with the properties of the underlying stellar population already covered in previous explorations. In addition, we deliver the full catalogue of H II regions,⁴ including all the spectroscopic properties extracted by our code. The code itself is also delivered for its use by the community.⁵

The layout of this article is as follows: In Section 2 we present the galaxy sample and data analysed in this study; the performed analysis is described in Section 3, including a summary of the procedure adopted to extract the information of the emission of stellar population and the emission lines (Section 3.1), and a detailed description of the procedure to detect and extract that information for the H II regions within each galaxy (Section 3.2). The final criteria adopted to select the H II regions are described in Section 3.3, and the proposed procedure to de-contaminate from the DIG emission is described in Section 3.4. The main results of our analysis are presented in Section 4, including (i) the main effects of the DIG decontamination (Section 4.1); (ii) the new proposed demarcation lines to distinguish H II regions from other sources of ionization (Section 4.2); (iii) the distribution of H II regions by morphology (Section 4.3) and (iv) by galactocentric distance (Section 4.4); and (v) the revisited relations between the properties of the underlying stellar population and the line ratios in H II regions (Section 4.5). Finally, we show the demarcation line for the other two most used diagnostic diagrams in Appendix A. We describe how to access the catalogue and the code in Appendix C. We show the equivalent width of H α , EW(H α), for the DIG sample in Appendix D.

2 SAMPLE AND DATA

We selected the analysed galaxies from the extended CALIFA (eCALIFA), the pilot CALIFA (pCALIFA), and PISCO samples (Sánchez et al. 2012a, 2016c; Galbany et al. 2018). eCALIFA comprises all galaxies observed by the CALIFA survey (Sánchez et al. 2012a), and all the extensions to the original sample observed using the same configuration and almost similar selection criteria (as described in Sánchez et al. 2016c). pCALIFA comprises the galaxies observed by the CALIFA survey in its pilot phase. PMAS/PPak Integral-field Supernova hosts COMpilation (PISCO; Galbany et al. 2018) sample comprises IFS data of 232 supernova host galaxies. The three surveys (eCALIFA, pCALIFA, and PISCO) were observed with the 3.5 m telescope at the Calar Alto Observatory (CAHA). eCALIFA + pCALIFA + PISCO comprises a homogeneous data set in terms of set up, depth, and redshift range. It includes galaxies selected in a similar way (basically matching the FoV of the instrument) observed along the last 9 yr without a fixed observing schedule. We based the current results on the 924 galaxies observed up to 2019 August. We will call them eCALIFA data set for simplicity along this article hereafter.

The details of the CALIFA survey, sample selection, observational strategy, and reduction are explained in Sánchez et al.

(2012a). The details of PISCO sample are explained in Galbany et al. (2018). The whole galaxy sample was observed using the PMAS spectrograph (Roth et al. 2005) in the PPAK configuration (Kelz et al. 2006). The instrument FoV has a hexagonal shape of $74 \text{ arcsec} \times 64 \text{ arcsec}$. In order to map the full optical extent of the galaxies up to $\sim 2.5R_e$ within this FoV, the galaxies of mother sample were selected by their diameter (Walcher et al. 2014), a primary criterion adopted for all extended sub-samples. Details of the particular selection criteria for each CALIFA (and PISCO) extension are given in Sánchez et al. (2016c). The complete coverage of the FoV is guaranteed with the observing strategy described in Sánchez et al. (2016c), with a final spatial resolution of FWHM $\sim 2.4 \text{ arcsec}$, corresponding to $\sim 0.8 \text{ kpc}$ at the average redshift of the survey. The spectroscopic resolution and sampled wavelength range for the V500 set-up (the currently adopted in this study) are $3745\text{--}7200 \text{ \AA}$ and $\lambda/\Delta\lambda \sim 850$, respectively. This resolution and range are sufficient to analyse the most important ionized gas emission lines from [O II] $\lambda 3727$ to [S II] $\lambda 6731$ in the redshift range of our galaxy sample and to deblend and subtract the underlying stellar population (e.g. Kehrig et al. 2012; Sánchez et al. 2012a; Cid Fernandes et al. 2013, 2014; Sánchez et al. 2016a).

The data were reduced using version 2.2 of the CALIFA pipeline. The modifications with respect to the early versions presented in Sánchez et al. (2012a), Husemann et al. (2013), and García-Benito et al. (2015) are described in detail in Sánchez et al. (2016c). In summary, the data fulfil the predicted quality-control requirements with a spectrophotometric accuracy that is better than 5 per cent everywhere within the explored wavelength range, both absolute and relative, with a depth that allows us to detect emission lines in individual H II regions as faint as $\sim 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ and with a signal-to-noise of $S/N \sim 3\text{--}5$. For the strong emission lines considered in this study, the S/N is well above this limit, and the measurement errors are negligible in most of the cases (as described in Sánchez et al. 2015a). Unavoidably, the S/N of the weak emission lines, e.g. [O III] $\lambda 4363$, could be less than $\sim 3\text{--}5$. In all cases, they have been propagated and included in the final error budget.

The final product of the data reduction is a regular-grid data cube, with two spatial dimensions (x and y coordinates corresponding to the right ascension and declination of the target, respectively) and one spectral dimension (z coordinate corresponding to a common step in wavelength). The CALIFA pipeline also provides a proper mask cube of bad pixels, the propagated error cube, and a prescription of how to handle the errors when performing spatial binning (due to covariance between adjacent pixels after image reconstruction). We use as the starting point of our analysis these data cubes, together with the ancillary data described in Walcher et al. (2014).

3 ANALYSIS

3.1 Stellar population and emission line properties

Spatially resolved stellar population and emission line properties provided by the PIPE3D pipeline (developed in purpose for the analysis of IFS data; Sánchez et al. 2016b) are used in this work. This pipeline has been extensively used in the study of CALIFA (e.g. Sánchez-Menguiano et al. 2016; López-Cobá et al. 2019; Méndez-Abreu, Sánchez & de Lorenzo-Cáceres 2019), MaNGA (e.g. Barrera-Ballesteros et al. 2016, 2017, 2018; Ibarra-Medel et al. 2016; Sánchez et al. 2018), SAMI (e.g. Sánchez et al. 2019), and MUSE (López-Cobá et al. 2017) data sets. PIPE3D adopts FIT3D as the core fitting package (Sánchez et al. 2016a), and the GSD156

⁴http://ifs.astrosu.unam.mx/CALIFA/HII_regions/

⁵<https://github.com/cespinosa/pyHIIexplorerV2>

simple stellar population library in the current implementation. This library that includes 156 templates: 39 stellar ages (from 1 Myr to 14.1 Gyr), and four metallicities ($Z/Z_{\odot} = 0.2, 0.4, 1, \text{ and } 1.5$) has been already used in many different previous publications (e.g. Pérez et al. 2013; González Delgado et al. 2014; Ibarra-Medel et al. 2016; Méndez-Abreu et al. 2019).

Details of the fitting procedure, the adopted dust attenuation curve, and limitations of the processing of the stellar populations are given in Sánchez et al. (2016a, b). We provide here a summary for understanding the nature of the properties explored in this study. First, a spatial binning is performed in each data cube to increase the S/N of the continuum ($S/N \sim 50$ per Å at 5000 Å) at any location within the FoV. This limit was selected as a compromise between not losing significant spatial information and providing accurate properties of the stellar populations. It is based on the simulations presented in Sánchez et al. (2016a). As shown in Sánchez et al. (2012a, 2016c), a significant fraction of the original spaxels (~ 50 percent) have an S/N above this goal, and those remain to unbind. For the remaining ones, the adjacent spaxels are aggregated, co-adding the corresponding spectra, to increase the S/N. However, contrary to other binning schemes (e.g. Voronoi binning, Cappellari & Copin 2003), the adopted one only aggregates those spaxels whose relative flux intensities differ less than 15 per cent from each other at the selected wavelength range (5000 Å). In this way, the original shape of the galaxy is better preserved, and there is no mixing of nearby regions that correspond to different structures (e.g. arm/inter-arms). The penalty is that the goal S/N is not reached for all the final bins, which in general, have an S/N well above 40 for the CALIFA data (see Ibarra-Medel et al. 2016, for some examples of the procedure). On average, ~ 1000 – 2000 tessellae (binned regions) are obtained, along with their corresponding spectra for each CALIFA data cube.

After the binning process, the stellar population is fitted for each of the binned spectra following two steps. As the first step, the stellar velocity and velocity dispersion, and the stellar dust attenuation, are derived using a limited sub-set of the stellar library described before. As the second step, a multi single stellar population (SSP) linear fitting is performed, using the full GSD156 library, adopting the kinematics and dust attenuation derived in the previous step. A Monte Carlo iteration is performed for the second step, perturbing the original spectra with their corresponding errors, to provide errors on the estimated stellar parameters. This procedure gives a stellar-population model for each tessella. Then, we derive a stellar-population model for each spaxel by re-scaling the model within each bin to the flux intensity of the continuum at the corresponding spaxel (within the considered bin; Cid Fernandes et al. 2013; Sánchez et al. 2016a). Using this model, we estimated the stellar population properties in each spaxel, generating maps of each property. The most relevant parameter considered in this article is the percentage of light and mass (weights) corresponding to each stellar population template within the library, in particular, for the fraction of young and old stellar populations (separated by a specific age limit, as we discuss later). Besides, several properties of the stellar component such as the stellar mass density, light-weighted (LW) and mass-weighted (MW) average stellar age ($\text{Age}_{\text{LW, MW}}$) and metallicity ($[Z/H]_{\text{LW, MW}}$), and the average dust attenuation (together with the kinematic properties already described before) are derived. The maps of each of these properties are packaged as channels (slices) of two data cubes (for convenience), comprising the average stellar properties (SSP cubes) and the weights of the decomposition of the stellar population (SFH cubes), as described in Sánchez et al. (2016b).

In order to derive the properties of the ionized gas emission lines, for each spaxel, the stellar spectrum model (as described above) is subtracted from the original one creating a ‘gas-pure cube’. This cube contains the emission lines together with the noise and the residuals of the stellar-population modelling. The main properties of a set of 51 emission lines are then derived for each spaxel based on a weighted momentum analysis (described in detail in Sánchez et al. 2016b), including for each line the integrated flux intensity, the velocity, velocity dispersion, and the EW. The set of emission lines comprises the most relevant ones detected in the optical spectrum of H II regions (see a complete list in Sánchez et al. 2007), as detected in the Orion nebulae using a similar wavelength range. The final data products of this analysis, described in Sánchez et al. (2016b, 2018), comprise a data cube for each galaxy in which each channel (slice) contains the spatial distribution (map) of each of the derived properties for each analysed emission line (plus the corresponding maps for their errors). These data products (labelled as `flux.elines` cubes), together with the ones storing the properties of the stellar populations, are the primary input for the further analyses performed in this article.⁶

3.2 H II detection and extraction

We performed the segregation of H II regions using PYHIIEXPLORER.⁷ This code is based on HIIEXPLORER, originally written in PERL (Sánchez et al. 2012b). PYHIIEXPLORER does essentially the same procedures of HIIEXPLORER, re-coding it in PYTHON, which is more commonly used nowadays in astronomy. This makes a easier to distribute, install, (semi-) automatic update, and a major compatibility with other PYTHON codes or packages. We optimized this new version to use all the advantages of data processing that PYTHON offers. Finally, we implement PYHIIEXPLORER to run in parallel so that it can process several galaxies at the same time, depending on the available number of cores.

The detection of ionized regions performed by PYHIIEXPLORER is based on two assumptions. (1) H II regions have strong emission lines that are clearly above the continuum emission and the average ionized gas emission across each galaxy. (2) The typical size of H II regions is about a few hundreds of parsecs, which corresponds to a usual projected size of a few arcsec at the distance of our galaxies. These assumptions will define clumpy structures with a high H α emission line contrast in comparison to the continuum.

Thus, the main input of the algorithm is an H α emission map (preferentially, as it is the most intense emission line in H II regions in the optical regime). The process to identify ionized regions consists of two steps: (i) the identification of the emission peaks (i.e. local maximum) from the input emission map, and (ii) the aggregation of adjacent pixels to each peak in order to segment the image. For this, the algorithm requires some additional input parameters: (i) a minimum flux intensity threshold for the peak emission; (ii) a minimum relative fraction with respect to the peak emission to keep aggregating nearby pixels; (iii) a maximum distance to the ionized region peak emission to stop the aggregation; and (iv) a minimum absolute intensity threshold in the adjacent pixels to continue the aggregation.

The first, second, and fourth parameters are intensity lower limits, and the third one is a physical-geometry parameter. The peak (local

⁶A subset of these products was distributed for a subsample of the current analysed galaxies in Sánchez et al. (2016b).

⁷<https://github.com/cespinosa/pyHIIexplorerV2>

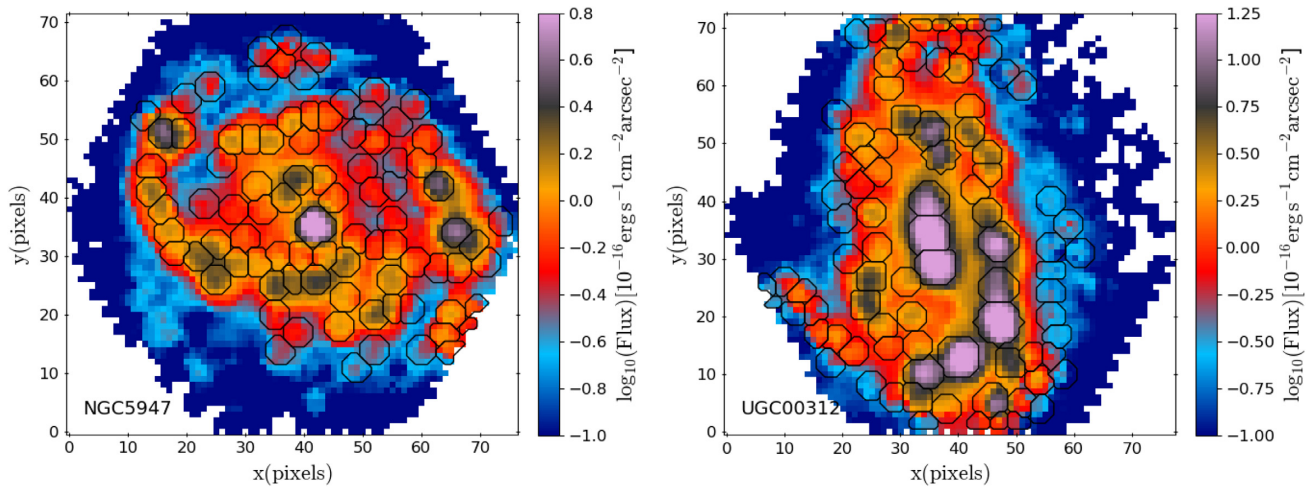


Figure 1. $H\alpha$ emission maps of NGC 5947 and UGC 00312 of the analysed sample; the colour code represents flux intensity. Black contours show the ionized regions found by PYHIIEXPLORER for each galaxy.

maximum) emission must be higher than the first parameter to define a new H II region. The algorithm starts looking for a peak (local maximum) and aggregates the adjacent pixels following the described criteria until no further one fulfils them. Then it looks for a new peak emission and iterates the process until no new local maximum is found. Finally, it iterates again over the segregation maps to redistribute pixels to the nearest peak when two or more adjacent H II regions overlap following the previous criteria.

All the parameters described above can be derived by visual inspection of the $H\alpha$ emission map or with a statistical analysis of the same emission maps (i.e. determining the 3σ intensity threshold, the typical $H\alpha$ flux of an H II region to our galaxy distance, and their typical size or the point-spread-function (PSF), resolution). In this particular implementation, for the PYTHON version, we adopted the same set of parameters introduced for the PERL one (Sánchez et al. 2012a). It allows us to use the same numerical values for parameters in both versions and make a direct comparison between the results.

The output of the algorithm is a FITS file that contains a segmentation map (see Fig. 1) similar in all ways to the one provided by other object detection algorithms like SEXTRACTOR (Bertin & Arnouts 1996). Each ionized region is identified by a unique ID value starting with 1. The pixels that are not associated with any ionized regions have a zero as ID. These areas are regions without emission, or their emission does not fulfil the criteria described before. Besides, a second FITS file is generated containing a masked map where all H II regions identified previously have been masked. The structures of both FITS files are described in Appendix C.

Finally, we extract the flux values of every emission line for each ionized region. For this, we calculate an $H\alpha$ luminosity-weighted mean with the fluxes in the original cubes (i.e. the data products of PIPE3D described in Section 3.1) for all spaxels with the same ID-index in the derived segmentation map. The final flux values and their associated errors are stored in tables for each ionized region ordered by rows (see Appendix C). Also, we extract the properties and the weights of the decomposition of the underlying stellar population (see Section 3.1) for each ionized region and store it in two other tables with similar structures as the flux tables (see Appendix C).

For this particular study, the input parameters used to identify the ionized regions from the $H\alpha$ emission maps of all galaxies analysed here are: (i) a flux intensity threshold for each ionized

region peak emission of $3 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-1}$, (ii) a minimum relative flux to the peak emission for associated spaxels corresponding to the same H II regions of 5 per cent, (iii) a maximum distance to the location of the peak of 5.5 arcsec, and (iv) an absolute flux intensity threshold of $0.5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-1}$ in the adjacent pixels to associate them to the peak emission. All these parameters were selected by a try-and-error procedure on a handful of galaxies, in order to maximize the number of detected H II without including clear diffuse ionized regions, based on the experience of previous studies (e.g. Sánchez et al. 2012b, 2015a; Sánchez-Menguiano et al. 2016, 2018). These parameters were derived by performing a visual inspection and a statistical analysis of the input emission line map, in order to maximize the number of detected ionized regions for the bulk sample of galaxies.

The current catalogue was generated by running PYHIIEXPLORER v2 on a computer system with an AMD® Ryzen® 2700x CPU with a 3.7 GHz clock speed. This CPU has 8 cores (and 16 threads) that can directly address 16 GB of RAM. The complete generation of the catalogue running PYHIIEXPLORER in parallel mode took ~ 25 min.

3.3 H II region catalogue

We apply the procedure outlined in the previous section over 924 data cubes corresponding to a similar number of galaxies observed within the eCALIFA + pCALIFA and PISCO survey. A total of 38 807 clumpy ionized regions were identified. In order of segregate the H II regions, from the clumpy ionized regions sample, we select those that fulfil the following criteria: (i) a minimum $S/N > 3$ in the $H\alpha$ flux; (ii) an $H\alpha/H\beta$ ratio above 2.7, to reject those regions with un-physical Balmer ratios; and finally (iii) a minimum $\text{EW}(H\alpha) = 6 \text{ \AA}$ and (iv) a fraction of young stars to the total luminosity (f_y) above of the 4 per cent was applied, following Sánchez et al. (2014).

The cuts in the $\text{EW}(H\alpha)$ and f_y were selected based on the previous knowledge of the properties of H II regions and after a detailed inspection of the distribution of both parameters (e.g. Sánchez et al. 2015a; Morisset et al. 2016; Lacerda et al. 2018). Fig. 2 shows the distribution of $\text{EW}(H\alpha)$ against the fraction of light in the V band of stars younger than 350 Myr (f_y) for (left-hand panel) all clumpy ionized regions selected by the code and (right-hand panel) all spaxels that do not belong to any clumpy ionized

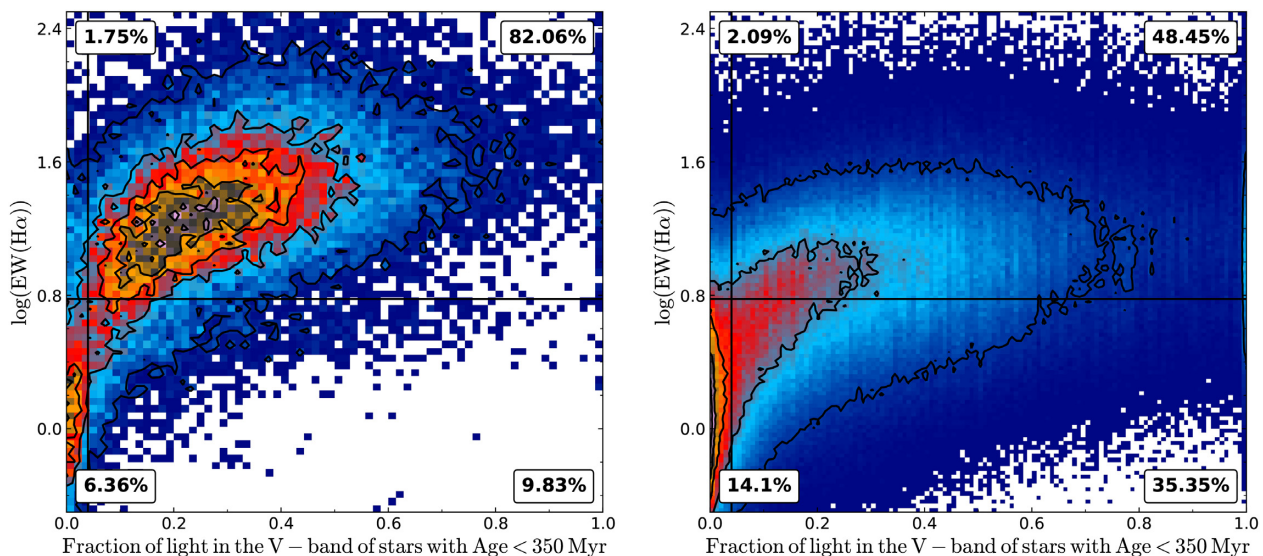


Figure 2. $\text{EW}(\text{H}\alpha)$ versus the fraction of light in the V band with age lower than 350 Myr for the underlying stellar population for the 38 807 detected ionized regions (*left-hand panel*) and for the pixels excluded by the segregation code (*right-hand panel*), i.e. the diffuse (not clumpy) regions. The light fractions are based on the SSP analysis performed by PIPE3D (Sánchez et al. 2016b). The vertical and horizontal lines (in both the panels) represent the adopted limits for the $\text{EW}(\text{H}\alpha)$ and fraction of young stars to identify H II regions. The contours (and colour code) show the density distribution of regions (or spaxels). The outermost contour encloses 95 per cent of the regions and each consecutive one encloses the 75, 55, and 35 per cent of the points. Finally, in both the panels we show the percentage of the number of ionized regions (*left-hand panel*) and DIG spaxels (*right-hand panel*) for each subdivision according our limits in $\text{EW}(\text{H}\alpha)$ and fraction of light.

region (i.e. not selected by the code). The vertical and horizontal lines correspond to the percentage of young stars equals 4 per cent, and the demarcation limits of $\text{EW}(\text{H}\alpha)=6 \text{ \AA}$, respectively. These limits were selected as a compromise between preserving the largest number of H II regions and removing the largest number of ionized regions that are not compatible with being an H II region (i.e. lack of enough young stars). Indeed, the actual limits correspond roughly to the location of the contour including 95 per cent of the clumpy regions with young stars (upper right) and with a lack of young stars (bottom left). Those regions that are identified as H II regions (upper right zone in the figure) represent the 67.7 per cent of the total clumpy ionized regions detected by the code.⁸

Applying all the criteria described above, we recover 26 408 H II regions. Unlike other selection methods to obtain H II regions commonly used in the literature, our method is based only on well-known properties of H II regions: a clumpy ionized gas structure with high $\text{H}\alpha$ emission and a young underlying stellar population. The first criterion is satisfied by construction; it is a fundamental part of the detection algorithm described in Section 3.2. The second one is provided by the adopted $\text{EW}(\text{H}\alpha)$ and f_y cuts that ensure that the ionized gas is indeed dominated by a young stellar population (f_y) and has a high contrast in the $\text{H}\alpha$ emission [$\text{EW}(\text{H}\alpha)$].

3.4 The diffuse ionized gas

It is well known that galaxies exhibit a low-intensity emission-line spectrum that it is broadly distributed along all their optical extension, known as DIG. Different authors have studied the DIG in the literature (e.g. Reynolds 1991; Rand 1996, 1998; Minter &

Balser 1997; Minter & Spangler 1997; Zhang et al. 2017; Lacerda et al. 2018). Unfortunately, its physical origin and properties are still not fully understood. It is detected far from the galactic plane as well as close to it (e.g. Flores-Fajardo et al. 2011, Levy et al. 2019), and in galaxies of any morphological type (e.g. Papaderos et al. 2013; Singh et al. 2013). The first detection of DIG was in the areas adjacent to classical H II regions on the Galactic disc (Reynolds 1971); however, later works proved the existence of DIG at large distances above the galaxy planes (e.g. Hoopes, Walterbos & Greenwalt 1996; Hoopes, Walterbos & Rand 1999). Oey et al. (2007) showed that DIG emission is present in galaxies of all types, and it may represent up to 60 per cent of the total $\text{H}\alpha$ flux intensity (Relaño, Peimbert & Beckman 2002). However, other authors reduce its contribution to ~ 4 –10 per cent when flux intensity is corrected by the dust attenuation (i.e. its contribution to the total luminosity; Sánchez et al. 2012b).

In the literature, the ionizing source of the DIG has been widely studied by different authors, too (e.g. Ferguson et al. 1996; Rand 1998). The radiation from massive OB stars leaked out from H II regions is one of the most widely accepted candidates (Haffner et al. 2009). Another one is ionization by hot low-mass evolved stars (HOLMES; Flores-Fajardo et al. 2011; Stasińska 2012) or post-AGBs stars (Papaderos et al. 2013). Finally, low-velocity shocks could contribute somehow to the DIG ionization (Dopita & Sutherland 1996; Monreal-Ibero, Arribas & Colina 2006; Monreal-Ibero et al. 2010), although its contribution is still not well estimated. Most probably, the ionizing source is different depending on the galaxy type, and on the location within galaxies (Lacerda et al. 2018; Sanchez 2019). In more early-type galaxies and in the retired areas of late-type ones, the two later ionizing sources are maybe the most common ones (e.g. Singh et al. 2013). On the other hand, in more late-type galaxies and in the vicinity of SF areas, the first proposed one is more likely to happens (Zurita, Rozas & Beckman 2000; Relaño et al. 2002; Zurita et al. 2002).

⁸In Appendix B, we compare the distribution of $[\text{O III}] \lambda 5007/\lambda 4959$ and the $[\text{N II}] \lambda 6584/\lambda 6548$ flux emission line ratios for the H II regions with the theoretical values.

As indicated before, DIG may be present anywhere in galaxies, and in particular, juxtaposed and mixed with the ionization of H II regions. This may contaminate the emission lines detected at that location, and alter the line ratios if the ionization is different than the one produced by nearby young massive stars themselves. The effects of DIG in low-resolution IFS data were explored by Mast et al. (2014) and Zhang et al. (2017), demonstrating that it can critically modify physical derivations like the oxygen abundance (recently addressed by Vale Asari et al. 2019). The physical spatial resolution of the eCALIFA data set is, on average, better than the resolution at which the spatial contamination effect is critical. However, these results highlight the need to attempt a correction of the DIG contamination in the emission line intensities derived for the detected H II regions.

Unfortunately, it is not possible to determine and decouple the contribution of the DIG in the H II regions from the individual spectra themselves: (i) The emissions by the H II regions are in general stronger, and it opaquely the emission of the DIG; (ii) at the spectral resolution of the data, both emissions have similar kinematic properties, precluding any decoupling. A tentative approach would be to estimate this contribution using the DIG emission at the adjacent areas of the detected H II regions and interpolate it. However, there are several potential problems in this approach. First, if the DIG is ionized by photons leaked from the H II regions themselves, then removing its contribution is unnecessary (at a first order), since it should show very similar emission properties, just re-scaled to lower intensity values. Secondly, due to the limited spatial resolution of our data, the wings of the PSF may create a considerable contribution to the adjacent areas. Again, this ionization corresponds to the H II regions and should not be removed. In order to decontaminate the DIG and preserve the contribution due to young stars, we characterize properties and determine the dominant ionizing source of DIG.

3.4.1 Characterizing the DIG ionizing source

A commonly adopted procedure to segregate the DIG ionized by old stars and the SF regions is using the $H\alpha$ surface brightness ($\Sigma_{H\alpha}$) since it is directly related to the density of ionized gas (Oey et al. 2007; Zhang et al. 2017). However, this criterion is valid only for face-on thin galaxies. Due to projection effects and for galaxies with any thickness, $\Sigma_{H\alpha}$ becomes an extensive quantity, since the light is co-added along the line of sight for the same observed area. This was recently discussed by Lacerda et al. (2018), who proposed a classification scheme based on the EW of $H\alpha$. This approach relies on the results by Cid Fernandes et al. (2010) and Sánchez et al. (2014), and on the fact that the EW($H\alpha$) is always an intensive quantity. According to them the ionized gas could be classified in three groups: (i) areas with $EW(H\alpha) < 3 \text{ \AA}$ correspond to diffuse gas being ionized predominantly by an old stellar population (HOLMES/post-AGBs); (ii) areas with $EW(H\alpha) > 14 \text{ \AA}$ correspond to the diffuse gas that is ionized by OB stars; and (iii) $3 \text{ \AA} < EW(H\alpha) < 14 \text{ \AA}$ corresponds to the diffuse gas that is powered by a mixture of both ionization sources.

Fig. 3 illustrates how the distribution of EW($H\alpha$) for the DIG regions (spaxels excluded from the clumpy regions by PYHIEXPLOER) correlates with respect to morphology (equivalent with the stellar ages). Early-type galaxies (NGC 6125, NGC 1349, and NGC 4003) tend to have low values of EW in comparison to late-type ones (NGC 5406, UGC 07012, and UGC 00312) (González Delgado et al. 2014). This is consistent with results obtained by

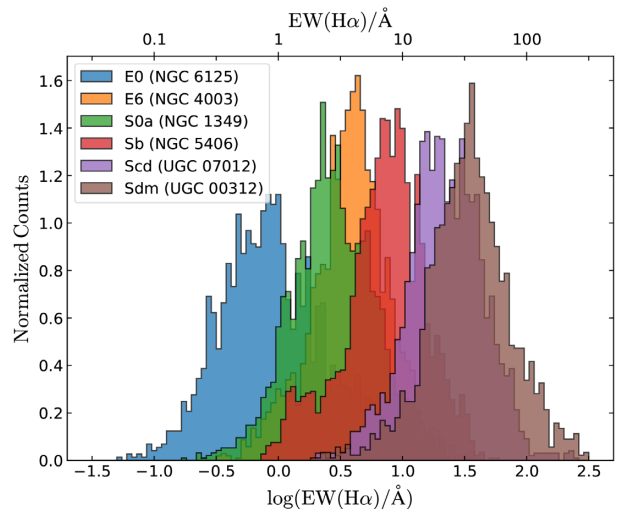


Figure 3. Normalized histograms of the EW($H\alpha$) of DIG spaxels along the Hubble sequence. One representative galaxy was chosen for six morphological types: E0 (NGC 6125), E6 (NGC 1349), S0a (NGC 4003), Scd (UGC 07012), Sb (NGC 5406), and Sdm (UGC 00312).

Lacerda et al. (2018). The late-type galaxies show equivalent widths that correspond to ionization due to old stars or leaking/PSF-wings contamination associated with SF areas. Therefore, for early-type galaxies, the associated ionizing sources are predominantly old stars as expected since the star formation is absent in these galaxies. Conversely, for late-type galaxies, both ionizing sources, old and young stars, should be observed. However, the number of spaxels with DIG associated with ionization by young stars should dominate over the number of them being ionized by old stars.⁹

In order to account for the effects of PSF wings, we repeat the selection of DIG areas by enlarging the masked areas around H II regions by one spaxel (i.e. 1 arcsec, ~ 700 pc at the redshift of the sample). However, although this procedure improves somehow the segregation, it does not guarantee a clean separation between the two ionizing sources. We repeat the procedure masking larger areas around the peak intensity of each H II region. The larger the radius, the smaller is the area selected as DIG in each galaxy. Furthermore, the number of galaxies with any area classified as diffuse decreases. Thus, the statistics become poorer. Following this procedure, we are not able to obtain a clear segregation between the different ionizing sources even for very large masking radius (3–4 times the PSF FWHM), when most of the area within the FoV of the data cubes is masked.

As a final approach, we explore the properties of the DIG gas of the sample, considering all galaxies together, trying to derive an average correction that would be applied to each individual spaxel in each data cube. For doing so, we select all the DIG spaxels again for all the galaxies as described in Section 3.4, i.e. by selecting those areas outside the segregation maps used to select the H II-region candidates. The green histogram in Fig. 4 shows the distribution of the EW($H\alpha$) of the diffuse selected in this way (raw DIG). The distribution shows a peak centred on $\sim 10 \text{ \AA}$, with an asymmetric shape, more tailed towards the low EW values. This distribution corresponds to that of the mixed DIG or intermediate ionization as described by Lacerda et al. (2018), showing a clear contamination

⁹ A further study about the DIG sample and their distribution across the BPT diagram is presented in Appendix D.

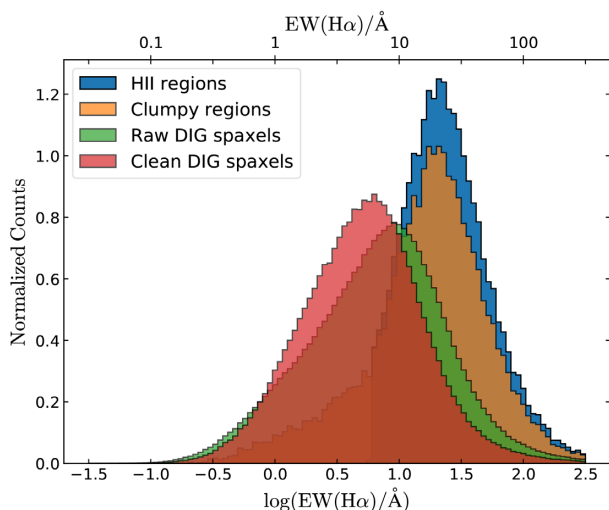


Figure 4. Normalized histograms of the $\text{EW}(\text{H}\alpha)$ for the four different sub-groups of ionized regions described in the text: (i) the raw DIG spaxels (DIG spaxels for all galaxies); (ii) clean DIG spaxels (DIG spaxels for the clean galaxy sample); (iii) clumpy ionized regions (all those regions identified by PYHIEXPLORER); and finally (iv) H II regions (clumpy ionized regions selected as H II regions; see Section 3.3).

by different sources of ionization. For comparison purposes, the orange histogram in the figure shows the distribution of $\text{EW}(\text{H}\alpha)$ for the clumpy ionized regions (as selected by PYHIEXPLORER). As expected, it is clearly peaked towards higher EWs ($\sim 30 \text{ \AA}$), although it presents a smooth but clear trend towards low EWs, so this indicates the presence of contamination by ionizing sources not compatible with H II regions.

In an attempt to select just the best suitable spectra to characterize the DIG, we start selecting only those galaxies with high-S/N diffuse emission. First, we exclude those galaxies in which (i) the mean value of half of the emission lines from all DIG spaxels has an SNR value lower than 1.5, and (ii) if any average value of the emission lines presents unrealistic values identified by eye for the DIG emission line fluxes (as a result of a bad fitting or a stellar population subtraction bias, or for regions at the edges of the hexagonal FoV of the data, in most of the cases). The remaining galaxies comprise 30 per cent of the galaxy sample (~ 300 galaxies, ~ 2.3 million of spaxels). Hereafter, we will refer to the spectra of this selection as the ‘clean DIG’ sample. The effects of this selection are clearly appreciated in the red histogram of Fig. 4. The distribution of $\text{EW}(\text{H}\alpha)$ for the ‘clean DIG’ sample has a peak now centred on $\sim 5 \text{ \AA}$, showing a more symmetric shape. The new values for the EWs are more in agreement with an ionization dominated by old stars. In order to investigate further the true nature of the ionizing source, we explore their distribution in the classical diagnostic diagrams.

Fig. 5 shows the $[\text{O III}]\lambda 5007/\text{H}\beta$ versus $[\text{N II}]\lambda 6583/\text{H}\alpha$ diagnostic diagram for all DIG spaxels. The spaxels were segregated by the morphology of their host galaxy in each panel, and the colour represents the mean $\text{EW}(\text{H}\alpha)$ at each location. The $\text{EW}(\text{H}\alpha)$ distribution for every group is plotted in an inset within each panel. There is a clear correlation between the location on the BPT diagram and the morphological type of the host galaxy. For early-type galaxies, the line ratios are mostly located in the intermediate area of the diagram (between the Kewley and Kauffmann curves). This region is dominated by a mixture of ionization by HOLMES/post-AGB and young stars. In particular, these line ratios cannot be

explained by SF models (Kewley et al. 2001, 2006; Kauffmann et al. 2003; Stasińska et al. 2006). The distribution for late-types galaxies is more located below the demarcation curves. This zone is dominated by regions that are classically ionized by young stars and leaking photons from H II regions.

Furthermore, the values of the $\text{EW}(\text{H}\alpha)$ depend on the morphological type as well. The DIG in early-type galaxies tends to have low values of $\text{EW}(\text{H}\alpha)$. Conversely, late-type galaxies show higher values. Therefore, the $\text{EW}(\text{H}\alpha)$ correlates with the morphology and with the position in the diagnostic diagrams. These results are expected by the correlation between the $\text{EW}(\text{H}\alpha)$ and the ionizing source, according to Lacerda et al. (2018). However, for low $\text{EW}(\text{H}\alpha)$ values, there are some spaxels below the Kauffmann’s demarcation line, i.e. they are in the area assumed to correspond to ionization by young stars or post-AGBs (Morisset et al. 2016). This effect is considerably more severe before cleaning the sample based on the criteria indicated before, as shown in Appendix D (Fig. D1), highlighting the need for the applied selection. In addition, there is a clear variation of the $\text{EW}(\text{H}\alpha)$ values and the location on the BPT diagram along the Hubble sequence, in agreement with Lacerda et al. (2018) and the examples are shown in Fig. 3.

We followed the same reasoning applied to select the H II regions from the clumpy ionized areas (Section 3.3) to select those spaxels within the clean DIG sub-set corresponding to ionization due to old stars adopting the inverse criteria. We recall that this criterion is based on the $\text{EW}(\text{H}\alpha)$ and f_y (see 3.3). Previously, we selected the regions with $\text{EW}(\text{H}\alpha) > 6 \text{ \AA}$ and $f_y > 4$ per cent as areas dominated by ionization due to young stars. On the contrary, to select DIG areas, we use the opposite constraints [$\text{EW}(\text{H}\alpha) < 6 \text{ \AA}$ and $f_y < 4$ per cent] as spaxels ionized by old stars. Fig. 2, right-hand panel, shows the distribution of $\text{EW}(\text{H}\alpha)$ against f_y for the clean DIG spaxels. A direct comparison between the distributions in this diagram for clumpy ionized regions (left-hand panel) and DIG ones (right-hand panel) shows that most of the clumpy regions are already associated with areas consistent with being ionized by young stars. In contrast, most of the DIG ones are concentrated in areas associated with ionization by old stars. We should note here that a fraction below 4 per cent of young stars is still reliable based on the inversion method adopted to derive the stellar population (FIT3D; Sánchez et al. 2016a). Different tests with similar codes indicate that below a 3 per cent the fraction of young stars is very unreliable (e.g. González Delgado et al. 2016), and below this limit, all fractions should be considered just upper limits (e.g. Bitsakis et al. 2019). On the other hand, based on our criteria, the upper right zone of the diagram for the DIG area corresponds, most probably, to contamination by leaking photons from H II regions and the contribution of the PSF wings (which have similar properties in this diagram). The two other remaining quadrants are much less populated, and they are most probably contaminated by low-S/N DIG areas, both in the continuum and the emission line properties, which shifts values from the other two ones that have a more clear physical interpretation.

3.4.2 Decontaminating the H II regions by the DIG

In the previous sections, we explore the distribution of the DIG and its nature. We selected a sub-set of galaxies from which it is possible to explore the properties of the DIG and define a procedure to determine which spaxels are ionized by old stars. In this section, we describe how we derive the average properties of the emission

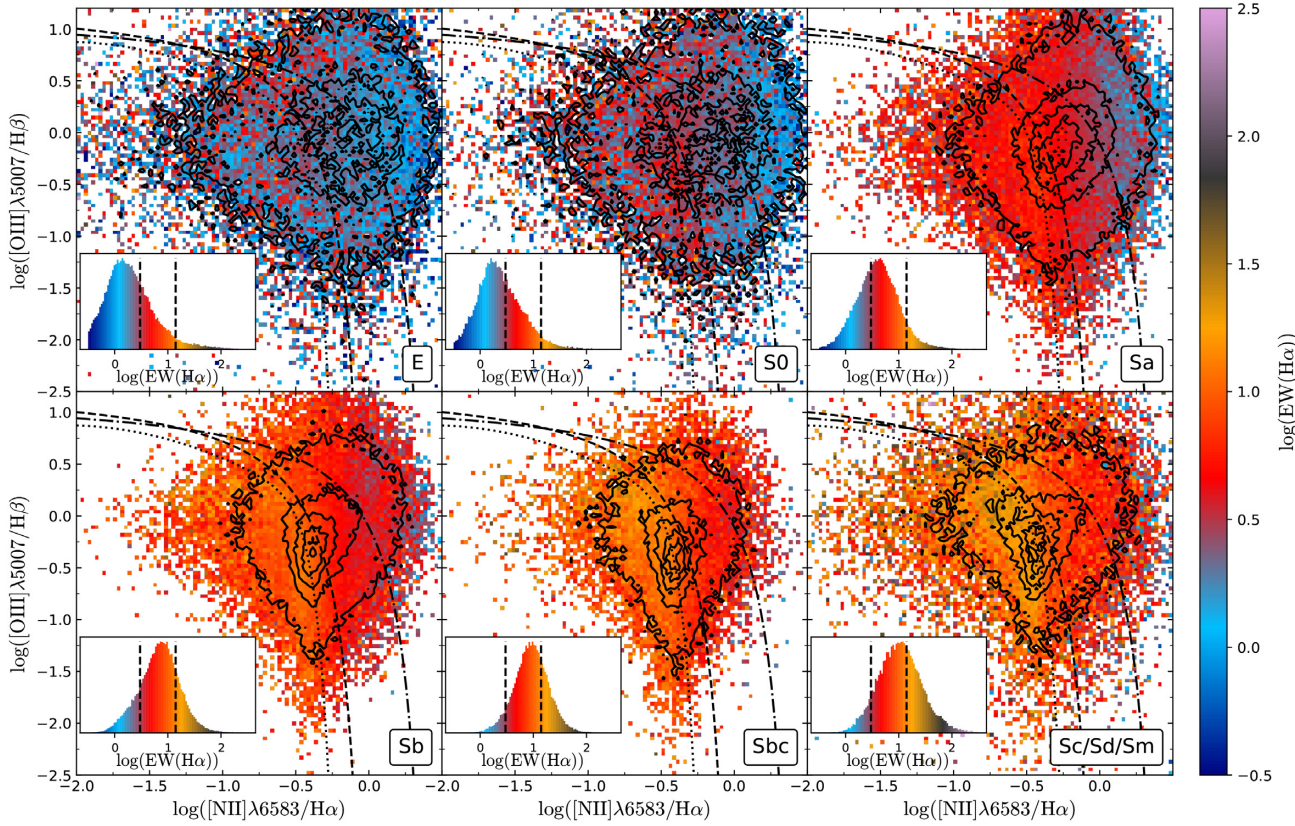


Figure 5. $[\text{OIII}]\lambda 5007/\text{H}\beta$ versus $[\text{NII}]\lambda 6583/\text{H}\alpha$ diagnostic diagrams for all the spaxels selected as *clean DIG*, segregated by the morphology of their respective host galaxies. The colour indicates the $\text{EW}(\text{H}\alpha)$. The inset within each panel shows the $\text{EW}(\text{H}\alpha)$ distribution of the corresponding spaxels. In these inset plots, we show the $\text{EW}(\text{H}\alpha)=3$ and 14 \AA as black dashed vertical lines. The dashed, dot-dashed, and dotted lines represent the Kauffmann et al. (2003), Kewley et al. (2001), and Stasińska et al. (2008) demarcation curves, respectively. The density contours are similar to those presented in Fig. 2.

line ratios for this final sub-set of spaxels (so far, for our data set, the best possible representation of the DIG). Then, we show how we correct the emission line intensities of our catalogue of H II regions from this DIG contamination.

For a j -th emission line flux of the i -th H II region, the corrected flux should be

$$F_{\text{corr}}^{i,j} = F_{\text{obs}}^{i,j} - F_{\text{diff}}^{i,j}, \quad (1)$$

where $F_{\text{obs}}^{i,j}$ is the observed emission line flux and $F_{\text{diff}}^{i,j}$ is the emission line flux of the DIG to be corrected from. Thus, we need to estimate this later flux intensity.

In order to do so, we estimate the average flux intensity of all DIG regions included in the *clean* sample. First, for each spaxel k -th within this sample of DIG areas we normalize the flux of each j -th emission line by the flux of $\text{H}\alpha$, defining the line ratio:

$$R_{\text{diff}}^{k,j} = \frac{F_{\text{diff}}^{k,j}}{F_{\text{diff}}^{k,\text{H}\alpha}}. \quad (2)$$

Then, we derive the error-weighted mean value of all these line ratios through all the *clean* sample of DIG spaxels, by using the formula

$$C_j = \frac{\sum_{k=1}^{k=n} w_{k,j} R_{\text{diff}}^{k,j}}{\sum_{k=1}^{k=n} w_{k,j}}, \quad (3)$$

where $w_{k,j}$ is the square of the inverse of the error of the j -th emission line for the k -th spaxel, and n is the total number of *clean* DIG spaxels.

Once derived C_j , i.e. the error-weighted mean value of the line ratio with respect to $\text{H}\alpha$ for the j -th emission line for the DIG, we can calculate the DIG emission line flux of that line that would correspond to a particular i -th H II region, $F_{\text{diff}}^{i,j}$, using the formula

$$F_{\text{diff}}^{i,j} = F_{\text{diff}}^{i,\text{H}\alpha} \times C_j, \quad (4)$$

where $F_{\text{diff}}^{i,\text{H}\alpha}$ is the $\text{H}\alpha$ flux of the diffuse gas. In principle, we do not have a direct way to estimate this flux. However, we can use the definition of the EW, and determine the relation

$$F_{\text{diff}}^{i,\text{H}\alpha} = \text{EW}_{\text{diff}}^{i,\text{H}\alpha} \times \rho^{i,\text{H}\alpha}, \quad (5)$$

where $\rho^{i,\text{H}\alpha}$ is the continuum flux density at the central wavelength of $\text{H}\alpha$. We have direct access to this flux density as a direct data product of the analysis by PIPE3D, which provides a version of the original data cube once we subtracted the estimated contribution of the emission lines. These data cubes are used to estimate the EWs of all the emission lines, as explained in Sánchez et al. (2016b). On the other hand, for the DIG (ionized by HOLMES or post-AGBs), we consider that $\text{EW}(\text{H}\alpha) \approx 1 \text{ \AA}$, based on theoretical estimations (Binette et al. 1994; Sarzi et al. 2010; Papaderos et al. 2013; Gomes et al. 2016b). Indeed, this value is in agreement with our results shown in the right-hand panel of Fig. 2. Although the observational mean value is $\sim 1.4 \text{ \AA}$, we prefer to use a conservative value, $\text{EW}(\text{H}\alpha) \approx 1 \text{ \AA}$, to perform the DIG correction. It is important

to note that this value is not related to the mean value shown in Fig. 4 since this distribution is dominated by the PSF-wings contamination and the photons leaking, as indicated before.

Using equations (5) and (4), and the parameters C_j estimated for each j -th emission line, we estimate $F_{\text{diff}}^{i,j}$. Finally, we decontaminate the contribution of the DIG for each emission line in each H II region within the catalogue using equation (1). The resulting values are stored in a new catalogue of H II regions corrected by DIG.

3.4.3 Caveats to the adopted procedure

We should note here that the adopted approach to derive the DIG contribution and its use to decontaminate line fluxes of H II regions is based on different assumptions described below that may not be entirely valid. We have assumed that the only DIG contribution to correct for is the one produced by old stars (HOLMES/post-AGBs), considering that the diffuse ionization by photon leaking would produce similar emission line ratios. That assumption is far from been fully proved (e.g. Relaño et al. 2012). Furthermore, Vale Asari et al. (2019) recently studied the effects of a decontamination by DIG ionized by photon leaking in the derivation of the oxygen abundance. Using SF galaxies from the MaNGA data set (Bundy et al. 2015), they found that there is a difference of up to 0.1 dex in the oxygen abundance calculated with the N2 index at the high metallicity end and low EW. However, there is no significant change in the oxygen abundances derived with other strong line indices (e.g. O23, O3N2, etc.). This agrees with our perception that this particular DIG does not alter reported line ratios significantly (besides maybe N2, with a small effect in any case).

On the other hand, other contributions, like shock ionization, that could be considerably important at different scales (Veilleux, Cecil & Bland-Hawthorn 2005; López-Cobá et al. 2019) are not considered here. Finally, we considered that all the DIG presents the same line ratios everywhere within a galaxy and for different galaxies. This is a very strong assumption that it is only valid at first order, as we have indeed shown in Fig. 5. The line ratios expected by HOLMES/post-AGB ionization cover a wide range of values (e.g. Gomes et al. 2016b; Morisset et al. 2016). However, our empirical exploration of these line ratios indicates a small degree of variation, almost compatible with the expected distribution due to their errors. In addition, we consider that $\text{EW}_{\text{diff}}^{H\alpha}$ is 1 Å everywhere across the optical extension of galaxies and for all galaxies. This assumption, based on theoretical expectations, is empirically verified for the DIG of the earlier-type galaxies (Fig. 5). However, for later morphological types, the observed EWs are slightly larger. We assumed that the contamination by other sources of DIG and the wings of the PSF (from the adjacent H II-region) alter the observed EWs of the DIG in these later-type galaxies. We acknowledge all these caveats to the adopted procedure. However, based on the performed experiments, we consider that this is the best approach we can adopt so far for the considered data set. Being aware of possible improvements in the procedure, we provide with an uncorrected version of the H II catalogue together with the corrected one to allow the community to perform their evaluation on the issue of the DIG contamination.

4 RESULTS

4.1 Effects of the DIG decontamination

Fig. 6 shows the distribution across the classical BPT diagnostic diagram for the final sample of H II regions before and after applying

the DIG correction (left-top and right-top panels). As can be seen, most of the uncorrected values are located below the maximum starburst envelope demarcation line (Kewley et al. 2001). This result is expected since the criteria to select H II regions are based on the physical properties of this type of nebulae. After DIG correction, there is a change in the emission line fluxes and, consequently, in the respective EWs, fluxes ratios, and the distribution across the diagnostic diagram. The distribution stretches at the edges, covering a slightly wider range of line ratios. However, the average trend is very similar to the contaminated fluxes. Most of the decontaminated values remain in the SF zone in the diagnostic diagram. Besides, the $\text{EW}(H\alpha)$ values decrease slightly on average, since the continuum flux density is constant along the DIG correction calculations (see Section 3.4.2), but the flux intensities have decreased slightly too. This effect is stronger for the H II regions with lower $\text{EW}(H\alpha)$.

This result is somewhat expected. Mast et al. (2014) have already shown the effects of the contamination by DIG on the flux ratios used in the explored diagnostic diagrams. Although our correction only takes into account a diffuse mainly powered by old stars and post-AGBs (among other assumptions), our results show the same trend that they obtained with their methodology. Unfortunately, the spatial resolution of our data does not allow us to identify all diffuse gas components accurately and, hence, our data are not fully decontaminated from the complete DIG contribution. However, the proposed correction is enough to produce a small but noticeable change in the distribution of H II regions along the BPT diagram.

Fig. 6 shows the classical BPT diagram for the H II regions sample corrected by DIG with (bottoms panels) a colour code representing the difference between the flux corrected and the flux uncorrected for each line ratio in the diagram.¹⁰ As can be seen, the most significant change in the emission line fluxes is in the edges of the distribution (upper left for $[\text{N II}]/H\alpha$ and lower right for $[\text{O III}]/H\beta$). These results are in agreement with the change of the density points between the original distribution and the decontaminated one (upper panels of the same figure). Also, the error-weighted mean value, C_j (see Section 3.4.2), for the corresponding line ratios is shown as a white star. The value of $\Delta[\text{N II}]/H\alpha$ and $\Delta[\text{O III}]/H\beta$ depends on the position of C_j .

4.2 An empirical demarcation line for H II regions

As indicated before, Fig. 6 shows the classic diagnostic diagram for the entire sample of H II regions (corrected and uncorrected by DIG). For comparison purposes, we include in Fig. 6 the demarcation lines defined by Kewley et al. (2001) (dashed-dotted line), Kauffmann et al. (2003) (dashed line), and Stasińska et al. (2006) (dotted line). Since each demarcation line was derived on different assumptions, they do not provide the same classification (what is appreciated in the figure).

The most conservative demarcation line, in terms of minimizing the possible contamination by other ionizing sources different from H II/SF regions, is the one proposed by Stasińska et al. (2006) (dotted line in Fig. 6). This curve, like the one proposed by Kewley et al. (2001) (dot-dashed line in Fig. 6), defines the maximum envelop of line ratios derived by the location of pure SF regions based on photoionization models. The observed differences between both curves are due to the different models adopted in their derivations. Finally, an empirical demarcation line was derived based on a large sample of SF SDSS galaxies (Kauffmann et al. 2003; in Fig. 6).

¹⁰Other diagnostic diagrams are shown in Appendix A.

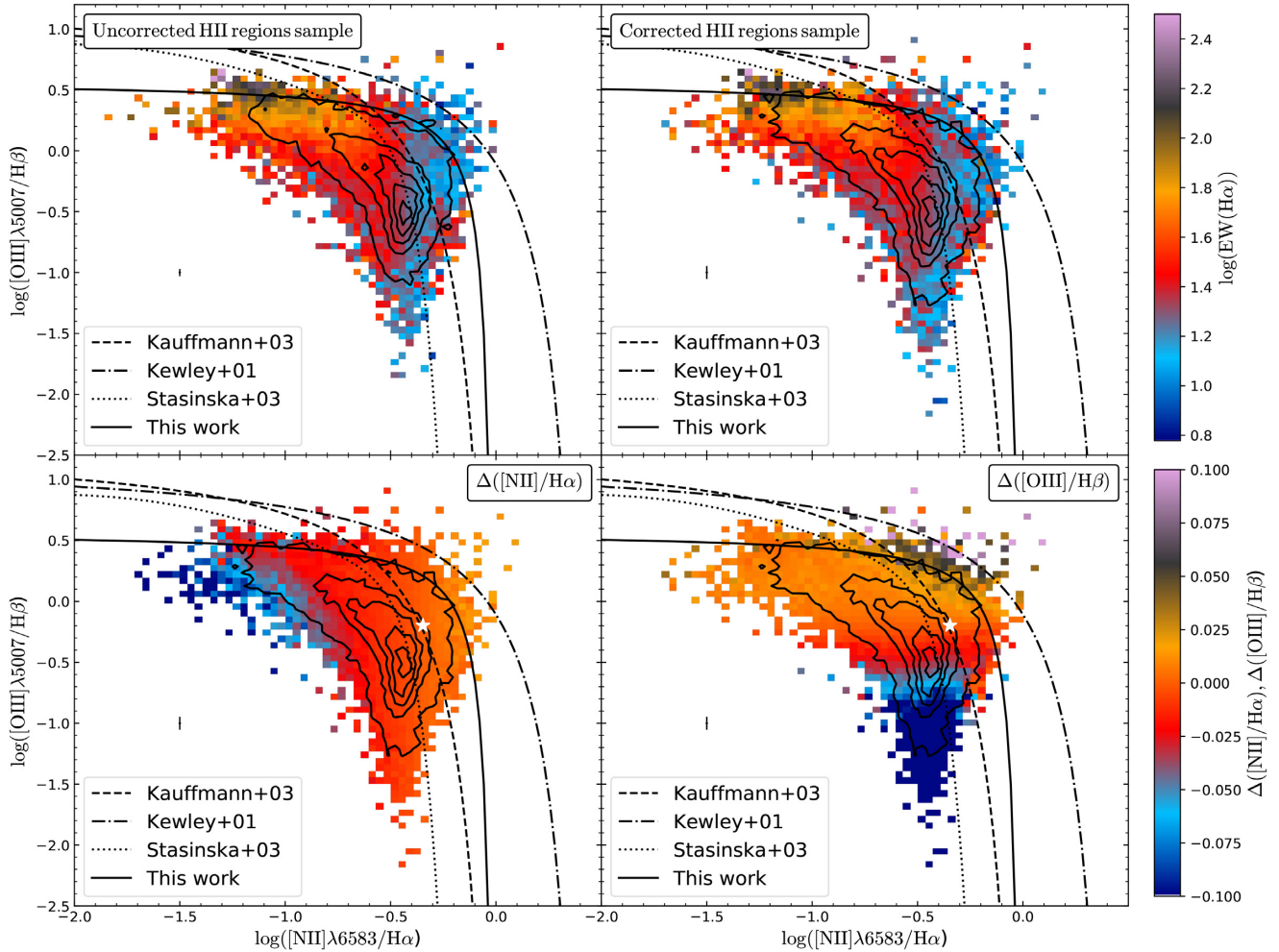


Figure 6. Distribution of the final sample of H II regions along the BPT diagram. *Left upper panel:* Distribution without the correction of the contamination of the DIG. *Right upper panel:* Distribution after correcting by the DIG contamination. In these panels, the colour code represents the EW of H α . *Left bottom panel:* Distribution of the corrected H II regions with a colour code representing the difference between the flux corrected and uncorrected for the [N II]/H α ratio (Δ [N II]/H α). *Right bottom panel:* Distribution of the corrected H II regions with a colour code representing the difference between the flux corrected and uncorrected for the [O III]/H β ratio (Δ [O III]/H β). In these panels, the white star is the error-weighted mean value for the corresponding line ratio. Notice that with the DIG correction the distribution has spread out slightly at the edges. This can be seen better on the density contours on the upper panel and in the colour gradient in the bottom panels. All lines represent the same demarcation curves described in Fig. 5. Also, the demarcation curve proposed in Section 4.2 is shown as a solid line.

This line was defined by inspecting the BPT diagram visually as the line that better distinguishes the AGN from the SF galaxies. We should advise the reader that none of these demarcation lines should be interpreted as strict boundaries between different ionization sources. As extensively discussed in previous articles (e.g. Sánchez et al. 2018, 2019), the demarcation lines should be interpreted as the upper limit of line ratios that can be achieved due to ionization by massive/young OB stars (Kewley et al. 2001). However, not all the ionizing regions located below that curve are associated with ionization by young stars per se. As clearly shown in Fig. 5, the DIG produced by old stars can be located below any of those curves, in particular, the one defined by Kewley et al. (2001). The regions between the Kewley and Kauffmann curves have been extensively classified as ‘mixed’ or ‘composite’. It is true that for single aperture spectroscopic data that encompass a large portion of the optical extension of galaxies or for low-resolution IFS data (and/or particular inclinations), these line ratios could be reproduced by the

combination of SF + AGN or SF + post-AGB ionizations (e.g. Mast et al. 2014; Davies et al. 2016). However, uncontaminated ionization by (i) certain types of H II regions, like the N-enhanced/shock mixed ones described by Ho et al. (2014) and Sánchez et al. (2012a); (ii) pure DIG produced by stars of certain ages (e.g. Morisset et al. 2016); (iii) pure low-velocity shocks (e.g. López-Cobá et al. 2019), and even (iv) low-metallicity AGNs (e.g. Stasińska 2017) may be located in that area without requiring a mixed ionization.

In this study, we propose a new purely empirical demarcation line that, like the other quoted ones, should be considered as an upper boundary for the involved line ratios due to ionization by young/massive stars. Based on the distribution of our sample of H II regions, we select the envelop that includes 95 per cent of them as the boundary of ionized regions mainly powered by young stars. Then, by fitting this set of line ratios that defines the limit using the same functional form adopted in the literature (e.g. Kewley et al.

2001; Kauffmann et al. 2003), we obtain the following demarcation lines for the most commonly used diagnostic diagrams:

$$\log([\text{O III}]/\text{H}\beta) < 0.13/(\log([\text{N II}]/\text{H}\alpha) - 0.003) + 0.57 \quad (6)$$

$$\log([\text{O III}]/\text{H}\beta) < 0.04/(\log([\text{S II}]/\text{H}\alpha) + 0.012) + 0.58 \quad (7)$$

$$\log([\text{O III}]/\text{H}\beta) < 0.056/(\log([\text{O I}]/\text{H}\alpha) + 0.40) + 0.61. \quad (8)$$

The curve for $[\text{O III}]/\text{H}\beta$ versus $[\text{N II}]/\text{H}\alpha$ (6) is represented by a solid line in Fig. 6. The locations of the proposed demarcation lines for the remaining classical diagnostic diagrams are shown in the corresponding figures in Appendix A (Figs A1 and A2).

It is important to note that these new demarcation lines are based on a sample of H II regions that were selected based on only two assumptions: (i) a clumpy structure with a high-contrast H α emission and (ii) an underlying stellar population comprising young stars (i.e. compatible with the assumed source of ionization).

4.3 H II regions along the Hubble sequence

Early explorations on the properties of H II regions have determined that later-type spirals present a larger number of H II regions, distributed along (more) open spiral arms than earlier-type ones (Kennicutt 1988, 1989). Furthermore, Kennicutt et al. (1989) and Ho et al. (1997) showed that H II regions in the central areas of early spirals distinguish themselves spectroscopically from those in the disc by their stronger low-ionization forbidden emission. These results were confirmed by more recent studies using large IFS data sets (Sánchez et al. 2012a, 2015b). In particular, Sánchez et al. (2015b) demonstrated that the ionization properties of H II regions strongly depend on the properties of the underlying stellar population, and, therefore, they present clear trends with the morphology as well. The enlarged H II catalogue presented here allows us to perform a more detailed exploration of this dependence.

Fig. 7 shows the distribution of H II regions across the BPT diagram segregated by the morphology of their respective host galaxies. The top left-hand panel shows the 254 regions found in E and S0 galaxies (~ 2 per cent of the total sample), corresponding to 38 galaxies. Although early-type galaxies are frequently associated with the absence of SF regions, recent results indicate that a small fraction of them present H II regions (e.g. Gomes et al. 2016a). They could be either remnants of a former disc or the result of a rejuvenation due to the capture of gas-rich galaxies or in falling gas (Gomes et al. 2016a). Table 1 shows the number of H II regions along the different morphological types. Most of the H II regions are located in late-type galaxies, with only ~ 7 percent of the H II regions hosted by E/S0 galaxies. However, most of the H II regions are not found in later spirals, but rather in Sb/Sbc galaxies, comprising 21/30 percent of the total sample. On average, the number of H II regions per galaxy increases from earlier to later types up to Sc, spanning from ~ 7 regions in E/S0 galaxies to ~ 50 in Sbc/Sc galaxies. Beyond that, for later spirals, the number decreases slightly. However, we cannot be completely sure if this decline is real or due to resolution effects. So far, the seminal studies by Kennicutt (1988, 1998) indicate that the number of H II regions should rise as later is the host, without any decline. The same articles indicate that H II regions in late-type galaxies are significantly brighter than those of more early types. However, the H II regions are limited by the properties of CALIFA's data. Inevitably, we are losing most of the low-luminosity H II regions. As we will discuss later (see Section 5), in a future work, we will explore the methodology used in this work on data with a higher spatial resolution.

The relation between the distribution of H II regions across the BPT diagram and the morphology of their host galaxy uncovered by Sánchez et al. (2015b) is appreciated in Fig. 7. The few H II regions found in early-type galaxies are located around the central region of the BPT diagram without any significant trend, maybe due to the low number of statistics. The trend becomes more evident when exploring the distributions from early-spirals (Sa) towards later ones, with H II regions shifting along the classical sequence of these objects (e.g. Osterbrock 1989).

The location in the BPT diagram of an H II region is primarily related to the ionization conditions of the nebulae (e.g. Kewley & Dopita 2002; Morisset et al. 2016). Thus, a priori, it is not expected a connection between this location and the morphological properties of the galaxy that hosts the H II regions, which is what we see precisely in Fig. 7. The low number of H II regions in early-type galaxies, in which SF activity is finished (or almost finished; Stasińska et al. 2008; Blanton & Moustakas 2009), was expected. However, the nature of the observed trend within the BPT diagram with the Hubble sequence, besides the clear increase in the number of H II regions, is less obvious. We can discard strong contamination by DIG as the primary driver for the observed distribution. First, the selection of clumpy ionized regions, together with the implemented cuts (Section 3.3), guarantees that we have indeed selected regions in which OB stars dominate ionization. Secondly, as shown in Section 4.1, this contamination shrinks somehow the range of covered parameters, but it does not blur the observed trends completely. Finally, we have performed a detailed DIG subtraction, so its contribution is minimized.

Assuming that OB stars indeed ionize all the selected regions, the observed trend should be the consequence of variations between these stars (and the surrounding nebulae). For the most late-type galaxies (Sd, Sm, and irregular), the location (upper left end on the diagram) is normally interpreted as the presence of low-metallicity stars with high ionization strengths. Meanwhile, for earlier-spirals (Sa or Sb), the location (lower right end on the diagram) is linked to higher metallicity stars with lower ionization strengths (these results are also found and discussed by Sánchez in press). Finally, in intermediate galaxy types, the H II regions extend within the two regimes. Despite the significant effect of the aging of the ionizing stars (which would impose certain variations within the observed trends), the only obvious conclusion is that we are observing a trend driven by a change in the metallicity, which induces a change in the ionization strength due to an anticorrelation¹¹ between both parameters (e.g. Dopita et al. 2016; Morisset et al. 2016; Pilyugin & Grebel 2016). If so, the observed trend implies a gradation with morphology in the stellar metallicity (in particular, the oxygen abundance).

4.4 H II regions along galactocentric distances

In the previous section, we explore the distribution of the H II regions segregated by the morphological type of their host galaxies across the diagnostic diagrams. In this section, we explore the distribution of the H II regions across the BPT for different galactocentric distances. The distance was directly derived from the H II region catalogue, which comprises the physical properties of the region by their location with respect to the centre of the galaxies, in addition

¹¹Recent studies found that in some galaxies there exists a positive correlation between both parameters (e.g. Poetrodjojo et al. 2018; Thomas et al. 2018).

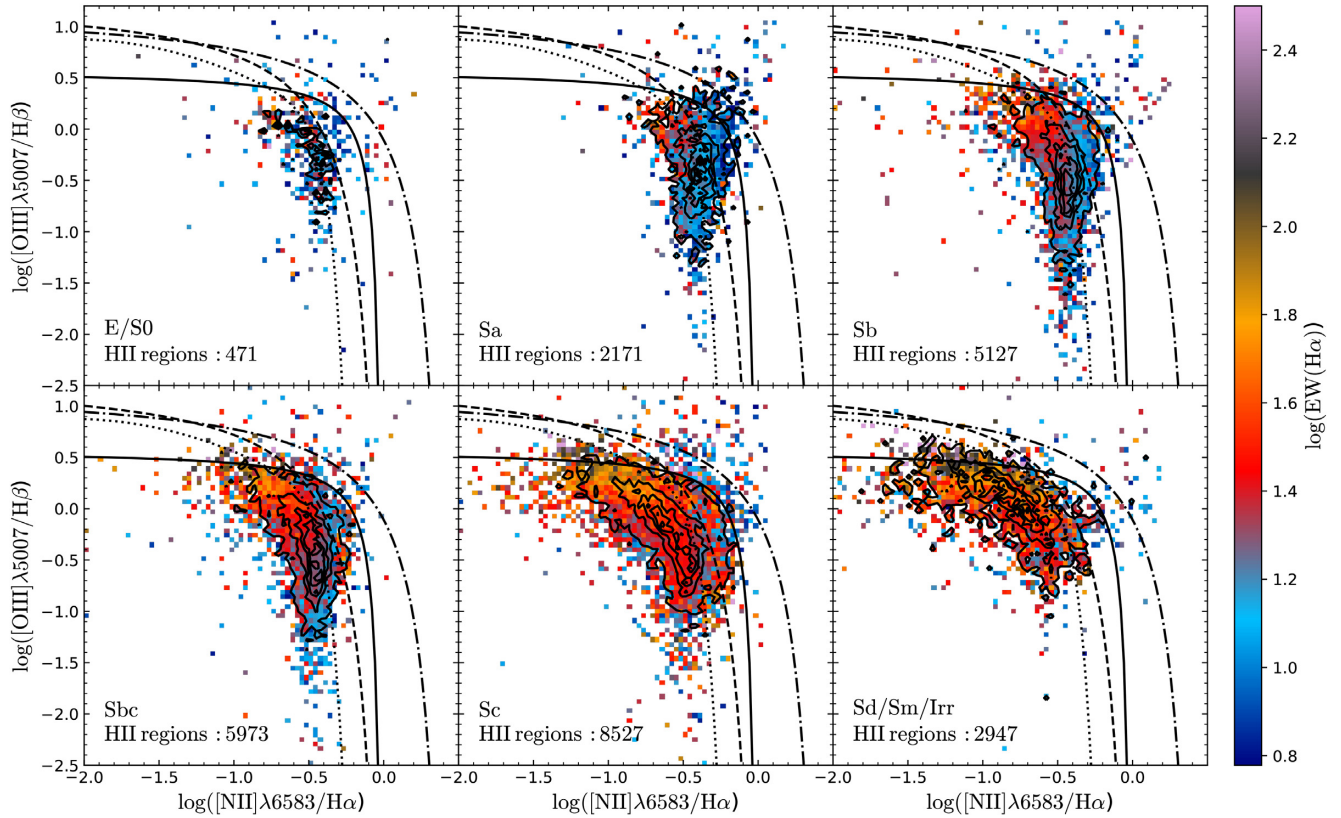


Figure 7. $[\text{O III}]\lambda 5007/\text{H}\beta$ versus $[\text{N II}]\lambda 6583/\text{H}\alpha$ diagram for the final sample of decontaminated H II regions. The sample was divided into six groups according to the morphological type of their host galaxy. The panels (from top-left to bottom-right) comprise the data for E/S0, Sa, Sb, Sbc, Sc, and Sd/Sm/Irr galaxies. In each panel, the colour code represents the mean $\text{EW}(\text{H}\alpha)$ of H II regions. The density contours are similar to those presented in Fig. 2. Also, the demarcation curve proposed in Section 4.2 is shown as a solid line.

Table 1. H II regions by morphology. (1) Morphology of the host galaxies. (2) Number of galaxies. (3) Number of galaxies with detected H II regions. (4) Fraction of galaxies with H II regions detected. (5) Number of H II regions per morphological type of host galaxy. (6) Fraction of H II regions. (7) Average number of H II regions per galaxy.

Morph.	N_{galT}	N_{gal}	$\%_{\text{gal}}$	$N_{\text{H II}}$	$\%_{\text{H II}}$	$\langle N_{\text{H II}} \rangle$
E	163	44	4.8	190	0.7	4.3
S0	105	46	5.0	415	1.5	9.0
Sa	135	121	13.1	2572	9.5	21.3
Sb	137	132	14.3	5598	20.6	42.4
Sbc	113	112	12.1	6274	23.1	56.0
Sc	172	172	18.6	8917	32.9	51.8
Sd	65	65	7.0	2259	8.3	34.8
Sm	21	21	2.3	633	2.3	30.1
Irr	10	10	1.1	214	0.8	21.4

to many other parameters. Based on these coordinates, we derive the galactocentric distance, correcting them by the inclination of each galaxy and position angle, and normalizing by the effective radius.

Fig. 8 shows the distribution of H II regions across the BPT diagram segregated into different galactocentric distances, ranging from the smallest to the largest distances. The distribution changes clearly with the distance to the centre of the galaxy. Those regions towards the centre ($R/R_e < 0.5$) are mostly located at the right end of the BPT diagram, at a very similar location where early-spirals H II regions are found (Fig. 7). On the other hand, the regions at larger distances ($R/R_e > 0.5$) follow the classical trend for these objects

(e.g. Osterbrock 1989), with a shift towards the upper left range of the distribution as the galactocentric distances increase. This result indicates that the ionization conditions of the H II regions depend on its location in the galaxy. This result is observed with the mean line ratios (white triangles) at different galactocentric distances. Earlier results already found that the classical H II regions are those within the disc of spiral galaxies (e.g. Diaz 1989). However, towards the centre of galaxies, they present stronger low-ionization forbidden emission lines (e.g. Kennicutt et al. 1989; Ho et al. 1997), shifting them towards the right-hand range of the diagram close to the demarcation curve (e.g. Sánchez et al. 2012a). These studies suggested that this may be due to the contamination by an additional

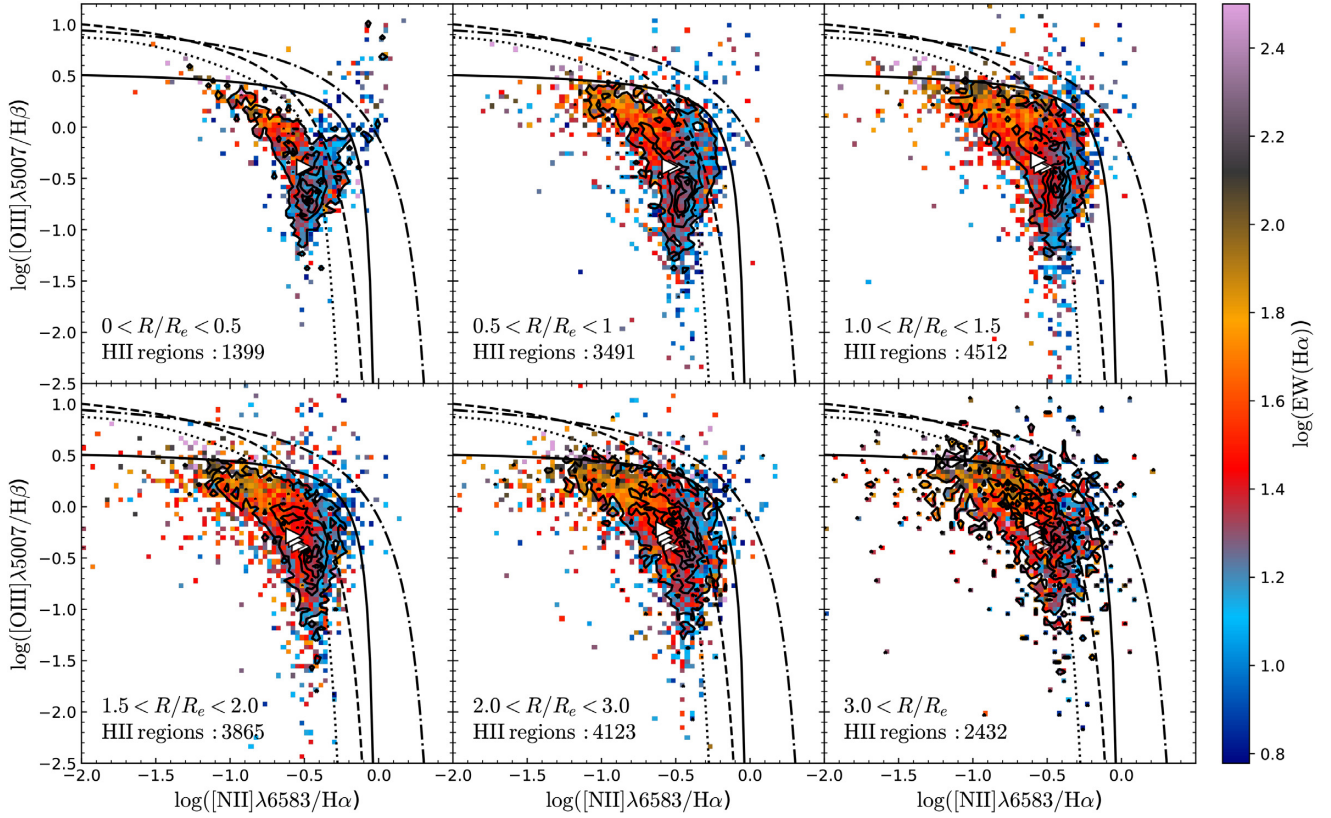


Figure 8. $[\text{O III}]\lambda 5007/\text{H}\beta$ versus $[\text{N II}]\lambda 6583/\text{H}\alpha$ diagram for the H II regions. The sample was divided into six groups according to their galactocentric distance, normalized to the effective radius: (i) $R/R_e < 0.5$ (top left hand); (ii) $0.5 < R/R_e < 1$ (top centre panel); (iii) $1 < R/R_e < 1.5$ (top right hand); (iv) $1.5 < R/R_e < 2$ (bottom left hand); (v) $2 < R/R_e < 3$ (bottom centre panel); and (vi) $R/R_e > 3$ (bottom right hand). The density contours are similar to those presented in Fig. 2. The colour code indicates the mean $\text{EW}(\text{H}\alpha)$ value in the respective bin. The white triangles in each panel represent the mean value of the $[\text{O III}]/\text{H}\beta$ and $[\text{N II}]/\text{H}\alpha$ ratios. Also, the previous values of each panel are shown in the next panels as transparent white triangles. All lines represent the same demarcation curves described in Fig. 5. Also, the demarcation curve proposed in Section 4.2 is shown as a solid line.

source ionization, as diffuse emission or shocks. However, the data in Fig. 8 are already corrected by the contribution of DIG and the distribution of H II regions remains at the same location in the BPT diagram. In general, we confirm the results by Sánchez et al. (2015b) in this regard, where the observed trends were already reported using a more limited number of H II regions (over a more limited number of galaxies) and without performing a DIG correction.

A plausible explanation for the enhanced low-ionization forbidden line ratios in central H II regions could be the contamination by shocks, as suggested by Ho et al. (1997). The presence of shock contamination in these regions was first suggested by Stauffer (1981) and explored to explain the differences in the disc and central H II regions by other authors (e.g. Peimbert, Torres-Peimbert & Ruiz 1992). Other authors do not report a clear difference between regions found at different galactocentric distances (e.g. Veilleux & Osterbrock 1987). Indeed, although there is a clear trend towards the right-hand regime of the BPT diagram for H II regions in the centre of galaxies, not all central H II regions are located at that extreme end. Therefore, if shocks are the source of the extra ionization, it is not present in all H II regions. A plausible explanation is that remnants of recent supernova explosions associated with the star formation events that ionized the observed region could contaminate the line ratios, deviating them from the expected location due to pure photoionization by OB stars. Under this scenario, our results indicate that central H II regions are more

prompt to this contamination. We will explore this possibility in forthcoming studies. As a first approach, models with shock (like MAPPINGS models; Sutherland & Dopita 2017) could be used to study the contribution of shocks, using, for example, the grid recently published by Alarie & Morisset (2019).

4.5 H II regions and the underlying stellar population

In the previous sections, we confirmed the results by Sánchez et al. (2015b), indicating that the location of an H II region within the BPT diagram is associated with the galactocentric distance and the morphology of the host galaxy. These authors connected the reported trends with the properties of the underlying stellar population (i.e. not the ionizing population) at the location (and host) of the H II regions. It is known that the bulk of the stellar population of the centre of disc galaxies presents similarities with that of early-type galaxies (e.g. González Delgado et al. 2014; Goddard et al. 2017; Sánchez et al. 2018). Moreover, they present radial gradients in both the stellar ages and metallicities (e.g. González Delgado et al. 2014), the sSFR (e.g. González Delgado et al. 2016; Sánchez et al. 2018) and the gas content (e.g. Ustom et al. 2017; Sánchez et al. 2018), and the overall star formation histories (García-Benito et al. 2017). Thus, the ionization conditions seem to be connected somehow with the local stellar evolution within the considered galaxy.

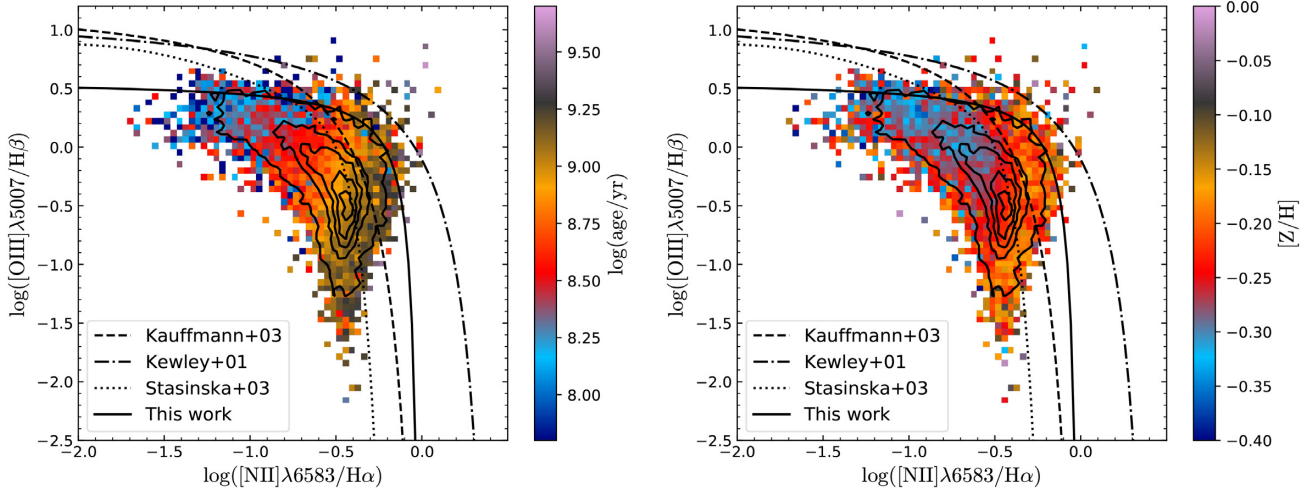


Figure 9. $[\text{O III}]\lambda 5007/\text{H}\beta$ versus $[\text{N II}]\lambda 6583/\text{H}\alpha$ diagram for the H II regions in our final catalogue. *Left-hand panel:* the colour indicates the average value of the luminosity-weighted age of the underlying stellar population. *Right-hand panel:* the colour indicates the average value of the luminosity-weighted metallicity of the underlying stellar population. The density contours are similar to those presented in Fig. 2. In both the panels, the dotted, dot-dashed, dashed, and solid lines are the demarcation curves used in Fig. 6.

Fig. 9 shows the distribution of H II regions across the BPT diagram with a colour code representing properties of underlying stellar populations. The left-hand panel shows the luminosity-weighted ages and the right-hand panel shows the metallicity of the underlying stellar population. The properties of the underlying stellar populations at the location of each H II region were extracted from the data products provided by PIPE3D, as described in Section 3.1, following Sánchez et al. (2013, 2014). As can be seen in Fig. 9, the location of an H II region in the BPT diagram depends on both stellar parameters. The regions with a younger underlying stellar population are located in the upper left area of the diagram. Otherwise, the regions with an older underlying stellar population are located in the right end of the diagram.

Similarly, the regions with metal-poor underlying stellar populations are found in the upper left area, while the regions with metal-rich underlying stellar populations are mostly located at the right end of the diagram. It is important to remark that we are not talking about properties of the ionizing population. The reported ages and metallicities correspond to the total stellar population at the location of each H II region. This result is consistent with the one obtained by Sánchez et al. (2015b), where similar trends were reported.

The trends along the BPT diagram with morphology and galactocentric distance reported in the previous sections are naturally explained by the distributions shown in Fig. 9. Early-type galaxies (and the centre of all galaxies) are dominated by old and metal-rich stellar populations (González Delgado et al. 2014; Sánchez et al. 2015a; Goddard et al. 2017). In general, and therefore, the H II regions on those galaxies are located at the right end of the distribution for these objects. On the contrary, late-type galaxies (and the outer areas of all galaxies) are mostly dominated by younger and more metal-poor stellar populations. Therefore, the H II regions on those galaxies are located in the left end of the distribution.

4.6 Imprints of the galactic evolution in the ionization

As described in the previous section, the location in the diagnostic diagrams depends on the underlying stellar population properties (age and metallicity). Thus, the BPT diagram exhibits a connection

between the stellar population properties and the line ratios. We explore this connection in here in a more quantitative way.

Fig. 10 shows the distribution of $[\text{N II}]/\text{H}\alpha$ line ratio (left-hand panel) and the $[\text{O III}]/\text{H}\beta$ line ratio (right-hand panel) along the luminosity-weighted age of the corresponding underlying stellar population at the location of each H II region. The colour code represents the stellar metallicity. As can be seen in the figure, there exists a trend between the line ratios and the stellar properties. Indeed, the distributions present a correlation coefficient of $r = 0.50$ (0.63) for low metallicities, $[\text{Z}/\text{H}] \sim -0.55$ dex, and $r = 0.46$ (0.48) for high metallicities, $[\text{Z}/\text{H}] \sim 0.15$ dex, for the $[\text{N II}]/\text{H}\alpha$ ($[\text{O III}]/\text{H}\beta$) line ratio. In general, the older and more metal-rich underlying stellar population is related to a higher (lower) value of the $[\text{N II}]/\text{H}\alpha$ ($[\text{O III}]/\text{H}\beta$) line ratio.

In order to illustrate more clearly the correlation described above, we averaged the values in bins of 0.1 dex in $[\text{Z}/\text{H}]$ and 0.25 dex in $\log(\text{age})$. The results are shown in Fig. 10, where every coloured circle is the mean value of the corresponding line ratio and age in each bin for a fixed metallicity (coded by the colour). The relation of the line ratios on both parameters is clear when explored the binned data.

To characterize the reported trends, we fit both line ratios with a linear combination of both properties of the stellar population properties (age and $[\text{Z}/\text{H}]$). We consider that with the current dispersion adopting a more complex functional form would not provide a better characterization. The best linear regression fitting for the $[\text{N II}]/\text{H}\alpha$ line ratio was

$$\log\left(\frac{[\text{N II}]}{\text{H}\alpha}\right) = 0.20_{\pm 0.02} \log(\text{age/yr}) + 0.06_{\pm 0.02} [\text{Z}/\text{H}] - 2.32_{\pm 0.26}. \quad (9)$$

The standard deviation of the residual of the line ratio, once we subtracted the best-fitting model, is ~ 0.166 dex. Compared with the standard deviation of the original distribution of values (~ 0.194 dex), the standard deviation was reduced¹² by a factor of ~ 52 per cent. Therefore, the $[\text{N II}]/\text{H}\alpha$ largely depends on the un-

¹²The difference was calculated by quadratic difference.

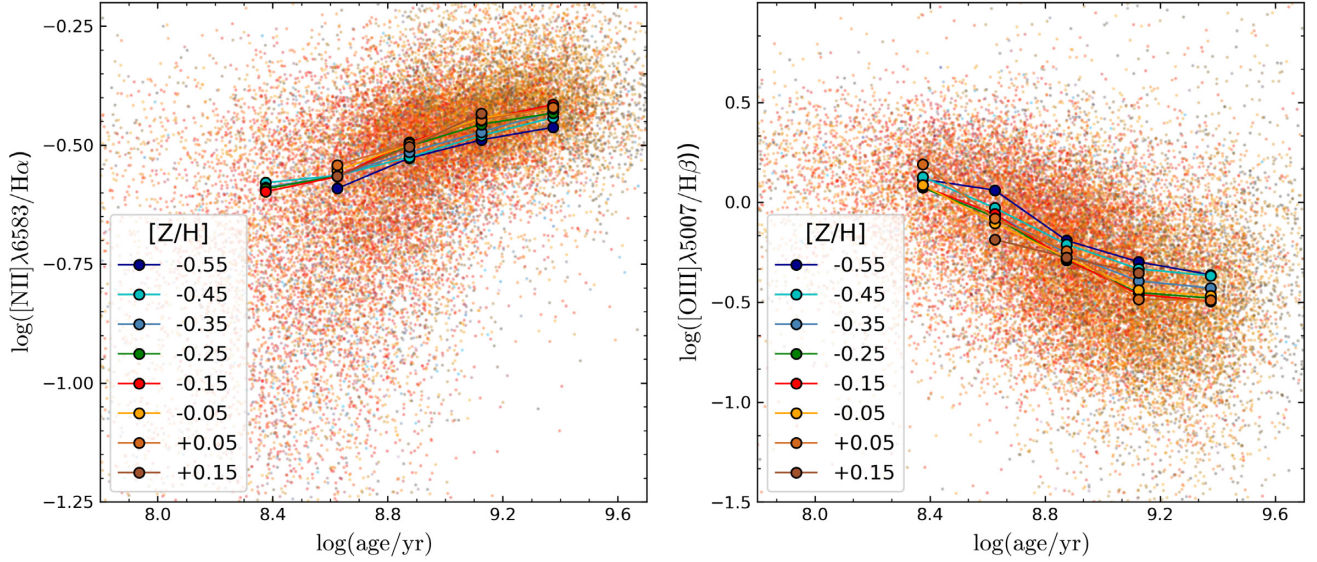


Figure 10. Distribution of the [N II]λ6583/Hα line ratio (left-hand panel) and [O III]λ5007/Hβ (right-hand panel) along the luminosity-weighted age of the underlying stellar population for the H II regions in our catalogue, colour coded by the [Z/H] of the underlying stellar population. The solid circles represent the average line ratios within bins of 0.25 dex in log(age/yr) for a fixed range of metallicities (binned in ranges of 0.1 dex). The average [Z/H] corresponding to each colour is included in the inset of each panel. All circles within the same metallicity range are linked by a solid line and plotted with the same colour (which codifies [Z/H]).

derlying stellar population. For [O III]/Hβ, the best linear regression fitting with the properties of the underlying stellar population is

$$\log\left(\frac{[\text{O III}]}{\text{H}\beta}\right) = -0.58_{\pm 0.01} \log(\text{age/yr}) - 0.17_{\pm 0.02} [\text{Z}/\text{H}] + 4.94_{\pm 0.12}. \quad (10)$$

The residual between the line ratio observed and the best fit presents a dispersion of ~ 0.348 dex. Again, compared with the initial dispersion of ~ 0.382 dex, the standard deviation was decreased by a factor of ~ 41 per cent. Like in the case of the previous line ratio, this modelling clearly illustrates the reported relation. However, despite the reduction of the dispersion, the line ratios still present a wide range of values. Other properties influence the observed line ratios besides the underlying stellar properties. In particular, the physical conditions of the nebulae (i.e. electronic density, ionization parameter, metallicity, the geometry of the ionized gas, etc.) must indeed influence the line ratios.

5 DISCUSSION AND CONCLUSIONS

In this study, we presented a new catalogue of H II regions extracted from the eCALIFA (+ pCALIFA) + PISCO sample. This new catalogue comprises information of $\sim 28\,000$ H II regions, including the flux of the most important emission lines in the optical range (between 3745 and 7200 Å) and the corresponding properties of the underlying stellar population. It is important to remark that we based the selection of H II regions on basically two assumptions: (i) they should appear as a clumpy structure in the Hα emission line map (see Section 3.2) and (ii) the underlying stellar populations should be compatible with the definition of H II region (see Section 3.3).

However, as can be seen in Fig. 1, the identification algorithm is not perfect. We are missing small clumpy regions and, also, we cannot resolve those regions that are too close one from each other. This is a direct consequence of the spatial resolution of the CALIFA data. As noticed by Mast et al. (2014), with a low spatial

resolution, the smallest ionized regions cannot be segregated and they are erroneously assigned to bigger adjacent ones or confused with the diffuse. This is why almost all identified ionized regions have a similar size in our catalogue. Thus, an important fraction of small ionized regions is lost in the final sample. This has a direct effect on the calculation of some properties, like the luminosity function (e.g. Sánchez et al. 2012b), that cannot be well recovered in the low-luminosity regime.

Furthermore, the wings of the PSF limit our ability to segregate among the different ionizing sources. First, they affect our ability to segregate between adjacent H II regions, as the boundaries of each clumpy ionized region are not well defined at the current coarse resolution. Therefore, the segregation algorithm developed to identify the ionized regions could erroneously associate pixels that contain certain emission coming from the PSF wing of any adjacent (brighter) region to a fainter one, modifying the estimated values of the derived parameters. Secondly, due to these wings, the DIG is polluted by the emission of any adjacent H II region too. Finally, H II regions could be contaminated by the emission of other possible ionizing sources (not only DIG but shocks or AGNs) as well. This may somehow affect the physical properties calculated from the emission line fluxes (Mast et al. 2014; Zhang et al. 2017).

To minimize the effects of this contamination, we present a new approach to correct the emission line fluxes of the H II regions by the contribution of the DIG. As we mention in Section 3.4, the DIG emission was calculated by considering all the spaxels in between those selected as being part of an H II region by our code. Again, spatial resolution plays an essential role in the proposed DIG correction. If we could identify exactly the limits of each ionized region (or model its shape), we would be able to identify clearly the DIG and, hence, make a proper correction irrespectively of the source of ionization of this diffuse (hot, old stars and leaking photons). Unfortunately, as indicated before and discussed in Section 3.4.3, we have an important contamination by ionization from the H II regions themselves due to the coarse resolution and

relatively large PSF wings. This contribution does not allow us to distinguish the DIG powered by leaking photons, for instance: The physical properties of this contamination, like $\text{EW}(\text{H}\alpha)$ and line ratios, are too similar to that of adjacent H II regions. Hence, we have corrected only the DIG powered by old, hot stars, which indeed is the one that may produce the strongest effects in the line ratios. This result is supported by the study of Vale Asari et al. (2019). They proposed a DIG correction where its contribution is calculated using only SF galaxies from the MaNGA data set, i.e. a DIG of higher $\text{EW}(\text{H}\alpha)$. There is a small but not significant change between the metallicities calculated by the corrected emission lines and the uncorrected ones for most of the explored calibrators (besides N2, which presents a 0.1 dex offset). Besides, the physical spatial resolution of the data analysed in Vale Asari et al. (2019) is nearly twice worse than the one included here (the FWHM ~ 0.8 kpc for CALIFA and ~ 2.5 kpc for MaNGA), which makes any correction by the DIG more complicated and with a stronger expected effect. To compare the contribution of all DIG types, we required observations with a high spatial resolution. However, the DIG correction present here, although not perfect, is an overall improvement with respect to not applying any correction at all.

It is clear that better spatial resolutions will facilitate the segregation of the clumpy ionized regions. It will improve the detection of smaller (thus, fainter) H II regions, the separation between adjacent regions, the definition of the DIG (and its different sources), and the overall decontamination of the DIG. New generation data sets, like the ones provided by MUSE (e.g. Sánchez-Menguiano et al. 2018) and SITELLE (e.g. Rousseau-Nepton et al. 2018), would require a re-evaluation of the proposed procedure, both for the detection of the H II regions and for the decontamination by the DIG.

A good example of the capabilities of new instruments is the study by Sánchez-Menguiano et al. (2018). They used MUSE observations of 102 galaxies, selected from the *All-weather MUSE Supernova Integral-field Nearby Galaxies* survey (AMUSING; Galbany et al. 2016; López-Cobá, submitted), to analyse the shape of oxygen abundance profiles. For doing so, they obtained a catalogue of H II regions using HIIEXPLORER. The total number of regions detected was 14 345. The spatial resolution of their data, FWHM ~ 0.8 arcsec (corresponding to the natural seeing) compared to the current value of 2.5 arcsec for CALIFA, allows them to identify a number equivalent to ~ 65 per cent of our current catalogue in a number of galaxies that comprises just ~ 10 per cent of the ones analysed here.

On the other hand, SITELLE provides observations where the natural seeing also limits resolution. Using this instrument, Rousseau-Nepton et al. (2018) identified 4285 H II region candidates in a single galaxy (NGC 628). However, the methodology used for the identification of H II regions generates several spurious ionized regions. In comparison, using the PINGS data (of similar projected resolution than CALIFA), and HIIEXPLORER, Sánchez et al. (2012b) recovered 373 H II regions in the same galaxy. Using the same data, we recovered 243 H II regions (from 244 ionized regions detected). The difference between the number of recovered H II regions by Sánchez et al. (2012b) and us is due to the input parameters used in HIIEXPLORER and PYHIIEXPLORER. In Sánchez et al. (2012b), the input parameters are listed.

In comparison, their minimum flux is lower than the value used in our calculations. However, they report that only 282 are H II regions with a good-quality spectrum and good estimation of the $\text{H}\beta$ flux. Meanwhile, the 243 H II regions that we report in here are the final number after applying the quality check test described in the text. Taking into account these details, we consider that the number of H II

regions found in both the studies is consistent and the differences are clearly justified. Thus, the use of SITELLE represents a theoretical improvement of a factor of 10 in the detection of these ionized sources.

Despite the limitations due to the coarse spatial resolution, there are still significant advantages in the current data set. MUSE does not cover the blue wavelength range of the optical spectrum below $\lambda < 4650$ Å. Therefore, fundamental emission lines are missing in the data provided by that instrument for galaxies at low redshifts. Moreover, the information provided for the stellar populations is limited since the critical wavelength range around the 4000 Å break is not covered. With SITELLE, the information for the stellar population is even less reliable due to the instrumental technique that affects the features of the continuum strongly. Finally, CALIFA allows us to present the current exploration for a statistically significant sample of galaxies, representative of the population at the nearby universe. Thus, the current H II region catalogue is not (a priori) biased by morphology, mass, or other galaxy properties. However, it is recognized that the possible properties derived from weak emission lines could drive to bias results, e.g. electron temperature derived from $[\text{O III}]\lambda 4363$. This is a direct effect of the inherent weakness of these emission lines and/or possible contamination from other sources (e.g. sky emission lines). However, the properties derived for weak emission lines could be calculated well enough from the H II regions.

We study the relationship between the location on the BPT diagram of the distribution of DIG-corrected H II regions and (i) the morphology of the host galaxies; (ii) the galactocentric distances of the H II regions; and (iii) the properties of the underlying stellar population. First, we demonstrate that the distribution of H II regions based on our simple selection criteria is located in the SF area of the diagnostic diagrams (without this being a prerequisite or part of the selection process). We found that the location in the BPT diagram depends strongly on the value of the $\text{EW}(\text{H}\alpha)$. This result was already reported in previous studies (e.g. Sánchez et al. 2015a; Lacerda et al. 2018). However, in comparison with them, we demonstrate that the result still holds when using a larger statistical sample of H II regions covering a more extensive range of physical properties and ionization conditions.

We clearly showed that the location in the diagnostic diagram depends on the morphological type of the host galaxy and the galactocentric distance. Early-type galaxies, in general, massive and metal rich, have H II regions located in the lower right end of the diagnostic diagram. Contrarily, in late-type galaxies, less massive and metal poor, the distribution moves towards the upper left zone of the diagrams. Thus, the change of location in the diagram is a consequence of the change of the stellar population properties along the Hubble sequence. González Delgado et al. (2015) already noticed that the metallicity, age, and other properties of stellar populations depend on the morphological type of galaxies. Indeed, they proved that the average stellar metallicity increases with the morphology.

Similarly, we found that the distribution in the BPT diagrams depends on the galactocentric distance. The most straightforward explanation is the same as the one described for the differences reported by morphological type. The centre of galaxies is populated by older and more metal-rich stellar populations, while the outer regions are populated by younger and more metal-poor ones (e.g. González Delgado et al. 2015; Sánchez et al. in press). This radial gradient is also evident in the oxygen abundance, which decreases with the galactocentric distances, too (e.g. Sánchez et al. 2012b, 2014).

This result is related to the inside-out scenario for the formation of galaxies. Other studies (e.g. Pérez et al. 2013; Ibarra-Medel et al. 2016; García-Benito et al. 2017) demonstrated the radial dependence on the star formation history and the time-scales of the gas inflow. Therefore, the chemical evolution of the ISM is affected by the star formation history at different distances resulting in the metallicity gradients on galaxies.

All the discussed results confirm that the locations within the BPT diagrams are tightly related to the properties of the underlying population, as already suggested by Sánchez et al. (2015a). This is clearly shown in Fig. 9. Moreover, we demonstrate that the line ratios $[\text{N II}]\lambda 6583/\text{H}\alpha$ and $[\text{O III}]\lambda 5007/\text{H}\beta$ present a clear correlation with both the ages and metallicities of the underlying stellar populations (Fig. 10). This relation is stronger for the $[\text{N II}]\lambda 6583/\text{H}\alpha$ line ratio. This result is somehow expected since this line ratio is widely used as a proxy of the oxygen abundance. The previous generation of stars enriched the chemical composition, oxygen, and other element abundances in the ionized regions. Hence, the oxygen abundance of the ISM naturally correlates with the metallicity of the younger stars, provided that the radial mixing is not strong (Sánchez et al. 2015a; Sánchez-Menguiano et al. 2016).

On the other hand, there is a correlation with the age of the stellar population due to the chemical enrichment processes in galaxies. It is well known that, in disc-dominated galaxies, stars are older and more metal rich in the central regions than the stars in the outer regions of the disc (Pérez et al. 2013; Sánchez-Blázquez et al. 2014; González Delgado et al. 2015). Thus, metallicity and its relation with the differential star formation history within galaxies and at different locations in them (e.g. García-Benito et al. 2017) are the most plausible explanation for the observed trends.

The other line ratio, $[\text{O III}]\lambda 5007/\text{H}\beta$, also shows a correlation with both analysed stellar parameters (age and $[\text{Z}/\text{H}]$). However, since this ratio is less sensitive to oxygen abundance and more sensitive to variations in the geometry, density, and temperature of the ionized nebulae, it presents a weaker correlation than that found for $[\text{N II}]\lambda 6583/\text{H}\alpha$.

Although other authors already noticed some of the results discussed above, we present them, for the first time, for a statistically well-defined sample of H II regions in the nearby Universe, several times larger than those analysed previously. We demonstrate that those results do not depend strongly on the possible contamination by DIG since we have applied a correction to that contribution. Moreover, we present a new demarcation line defined with this sample that is consistent with well-known previous ones, although it is totally unbiased regarding the involved line ratios.

The current catalogue of H II regions and the full set of emission lines discussed in this article are publicly available. To our knowledge, this is the largest catalogue of H II region properties, derived from the broadest range of galaxies of different morphologies and masses. With the analysis presented in this study, we have demonstrated that this sample is suitable to perform further explorations on the physical properties of these nebulae (in particular, for those that do not require detailed knowledge of the size and absolute luminosity). In forthcoming studies, we will present the main patterns of physical properties for these H II regions.

ACKNOWLEDGEMENTS

We thank the referee for the helpful comments that have improved the quality of this article. We are grateful for the support of CONA-

CYT grants CB-285080, FC-2016-01-1916, and CB2015-254132, and funding from the DGAPA/PAPIIT-IA101217, DGAPA/PAPIIT-IN100519, and DGAPA/PAPIIT-107215 (UNAM) projects.

This study uses data provided by the CALIFA survey (<http://califa.caha.es/>); it is based on observations collected at the Centro Astronómico Hispano Alemán (CAHA) at Calar Alto, operated jointly by the Max-Planck-Institut für Astronomie and the Instituto de Astrofísica de Andalucía (CSIC).

LG was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 839090.

This research makes use of python (<http://www.python.org>), of Matplotlib (a framework to the creation of scientific plots for Python; Hunter 2007), Astropy (a Python package for astronomy; Astropy Collaboration 2013; Price-Whelan et al. 2018), and SciPy (a Python-based ecosystem of open-source software for mathematics, science, and engineering; Jones et al. 2001).

REFERENCES

- Alarie A., Morisset C., 2019, *Rev. Mex. Astron. Astrofis.*, 55, 377
 Anderson L. D., 2014, *Am. Astron. Soc. Meeting Abstr.*, 223, 312.01
 Astropy Collaboration, 2013, *A&A*, 558, A33
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5 (BPT)
 Barrera-Ballesteros J. K. et al., 2016, *MNRAS*, 463, 2513
 Barrera-Ballesteros J. K. et al., 2018, *ApJ*, 852, 74
 Barrera-Ballesteros J. K., Sánchez S. F., Heckman T., Blanc G. A., The MaNGA Team, 2017, *ApJ*, 844, 80
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Binette L., Magris C. G., Stasińska G., Bruzual A. G., 1994, *A&A*, 292, 13
 Binette L., Flores-Fajardo N., Raga A. C., Drissen L., Morisset C., 2009, *ApJ*, 695, 552
 Bitsakis T. et al., 2019, *MNRAS*, 483, 370
 Blanton M. R., Moustakas J., 2009, *ARA&A*, 47, 159
 Bresolin F., 2019, *MNRAS*, 488, 3826
 Bundy K. et al., 2015, *ApJ*, 798, 7
 Cairós L. M., Caon N., García Lorenzo B., Kelz A., Roth M., Papaderos P., Streicher O., 2012, *A&A*, 547, A24
 Cappellari M., Copin Y., 2003, *MNRAS*, 342, 345
 Cid Fernandes R. et al., 2013, *A&A*, 557, A86
 Cid Fernandes R. et al., 2014, *A&A*, 561, A130
 Cid Fernandes R., Stasińska G., Schlickmann M. S., Mateus A., Vale Asari N., Schoenell W., Sodré L., 2010, *MNRAS*, 403, 1036
 Croom S. M. et al., 2012, *MNRAS*, 421, 872
 Davies R. L. et al., 2016, *MNRAS*, 462, 1616
 Diaz A. I., 1989, in Beckman J. E., Pagel B. E. J., eds, *Evolutionary Phenomena in Galaxies*. Cambridge Univ. Press, Cambridge, p. 377
 Dopita M. A., Sutherland R. S., 1996, *ApJS*, 102, 161
 Dopita M. A., Kewley L. J., Heisler C. A., Sutherland R. S., 2000, *ApJ*, 542, 224
 Dopita M. A., Rich J., Vogt F. P. A., Kewley L. J., Ho I.-T., Basurah H. M., Ali A., Amer M. A., 2014, *Ap&SS*, 350, 741
 Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, *Ap&SS*, 361, 61
 Dottori H. A., 1987, *Rev. Mex. Astron. Astrofis.*, 14, 463
 Dottori H. A., Copetti M. V. F., 1989, *Rev. Mex. Astron. Astrofis.*, 18, 115
 Ferguson A. M. N., Wyse R. F. G., Gallagher J. S. III, Hunter D. A., 1996, *AJ*, 111, 2265
 Flores-Fajardo N., Morisset C., Stasińska G., Binette L., 2011, *MNRAS*, 415, 2182
 Galbany L. et al., 2016, *MNRAS*, 455, 4087
 Galbany L. et al., 2018, *ApJ*, 855, 107
 García-Benito R. et al., 2015, *A&A*, 576, A135
 García-Benito R. et al., 2017, *A&A*, 608, A27
 García-Benito R., Pérez E., Díaz Á. I., Maíz Apellániz J., Cerviño M., 2011, *AJ*, 141, 126

- Garnett D. R., 2002, *ApJ*, 581, 1019
- Goddard D. et al., 2017, *MNRAS*, 466, 4731
- Gomes J. M. et al., 2016a, *A&A*, 585, A92
- Gomes J. M. et al., 2016b, *A&A*, 588, A68
- González Delgado R. M. et al., 2014, *A&A*, 562, A47
- González Delgado R. M. et al., 2015, *A&A*, 581, A103
- González Delgado R. M. et al., 2016, *A&A*, 590, A44
- Haffner L. M. et al., 2009, *Rev. Mod. Phys.*, 81, 969
- Hodge P. W., Kennicutt R. C. Jr, 1983, *ApJ*, 267, 563
- Ho I.-T. et al., 2014, *MNRAS*, 444, 3894
- Hoopes C. G., Walterbos R. A. M., Greenwalt B. E., 1996, *AJ*, 112, 1429
- Hoopes C. G., Walterbos R. A. M., Rand R. J., 1999, *ApJ*, 522, 669
- Ho L. C., Filippenko A. V., Sargent W. L. W., 1997, *ApJ*, 487, 579
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Husemann B. et al., 2013, *A&A*, 549, A87
- Ibarra-Medel H. J. et al., 2016, *MNRAS*, 463, 2799
- Jones E. et al., 2001, SciPy: Open Source Scientific Tools for Python. <http://www.scipy.org/>
- Kauffmann G. et al., 2003, *MNRAS*, 346, 1055
- Kehrig C. et al., 2012, *A&A*, 540, A11
- Kehrig C., Vílchez J. M., Sánchez S. F., Telles E., Pérez-Montero E., Martín-Gordón D., 2008, *A&A*, 477, 813
- Kelz A. et al., 2006, *PASP*, 118, 129
- Kennicutt Robert C. J., 1998, *ARA&A*, 36, 189
- Kennicutt R. C. Jr, 1988, *ApJ*, 334, 144
- Kennicutt R. C. Jr, 1989, *ApJ*, 344, 685
- Kennicutt R. C. Jr, Keel W. C., Blaha C. A., 1989, *AJ*, 97, 1022
- Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35
- Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, 556, 121
- Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, *MNRAS*, 372, 961
- Knapen J. H., 1998, *MNRAS*, 297, 255
- Lacerda E. A. D. et al., 2018, *MNRAS*, 474, 3727
- Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, *A&A*, 80, 155
- Levy R. et al., 2019, *ApJ*, 882, 84
- López-Cobá C. et al., 2017, *ApJ*, 850, L17
- López-Cobá C., Sánchez S. F., Bland-Hawthorn J., Moiseev A. V., Cruz-González I., García-Benito R., Barrera-Ballesteros J. K., Galbany L., 2019, *MNRAS*, 482, 4032
- López-Sánchez A. R., Esteban C., 2009, *A&A*, 508, 615
- Marino R. A. et al., 2013, *A&A*, 559, A114
- Marino R. A. et al., 2016, *A&A*, 585, A47
- Mármol-Queralto E. et al., 2011, *A&A*, 534, A8
- Martin P., Roy J.-R., 1994, *ApJ*, 424, 599
- Mast D. et al., 2014, *A&A*, 561, A129
- Méndez-Abreu J., Sánchez S. F., de Lorenzo-Cáceres A., 2019, *MNRAS*, 484, 4298
- Minter A. H., Balser D. S., 1997, *ApJ*, 484, L133
- Minter A. H., Spangler S. R., 1997, *ApJ*, 485, 182
- Monreal-Ibero A., Arribas S., Colina L., 2006, *ApJ*, 637, 138
- Monreal-Ibero A., Arribas S., Colina L., Rodríguez-Zaurín J., Alonso-Herrero A., García-Marín M., 2010, *A&A*, 517, A28
- Morisset C., Georgiev L., 2009, *A&A*, 507, 1517
- Morisset C. et al., 2016, *A&A*, 594, A37
- Moustakas J., Kennicutt R. C. Jr, 2006, *ApJS*, 164, 81
- Oey M. S., Parker J. S., Mikles V. J., Zhang X., 2003, *AJ*, 126, 2317
- Oey M. S. et al., 2007, *ApJ*, 661, 801
- Osterbrock D. E., 1989, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*. Univ. Sci. Books, Herndon, VA, USA
- Papaderos P. et al., 2013, *A&A*, 555, L1
- Peimbert M., Torres-Peimbert S., Ruiz M. T., 1992, *Rev. Mex. Astron. Astrofis.*, 24, 155
- Pérez E. et al., 2013, *ApJ*, 764, L1
- Pilyugin L. S., Grebel E. K., 2016, *MNRAS*, 457, 3678
- Poetrodjojo H. et al., 2018, *MNRAS*, 479, 5235
- Price-Whelan A. M. et al., 2018, *AJ*, 156, 123
- Rand R. J., 1996, *ApJ*, 462, 712
- Rand R. J., 1998, *ApJ*, 501, 137
- Relaño M., Peimbert M., Beckman J., 2002, *ApJ*, 564, 704
- Relaño M., Kennicutt R. C. Jr, Eldridge J. J., Lee J. C., Verley S., 2012, *MNRAS*, 423, 2933
- Reynolds R. J., 1971, PhD thesis. Univ. Wisconsin - Madison
- Reynolds R. J., 1991, in Bloemen H., ed., *Proc. IAU Symp. 144, The Interstellar Disk-Halo Connection in Galaxies*. Kluwer, Dordrecht, p. 67
- Rosales-Ortega F. F., Kennicutt R. C., Sánchez S. F., Díaz A. I., Pasquali A., Johnson B. D., Hao C. N., 2010, *MNRAS*, 405, 735
- Rosales-Ortega F. F., Díaz A. I., Kennicutt R. C., Sánchez S. F., 2011, *MNRAS*, 415, 2439
- Roth M. M. et al., 2005, *PASP*, 117, 620
- Rousseau-Nepton L., Robert C., Martin R. P., Drissen L., Martin T., 2018, *MNRAS*, 477, 4152
- Sánchez-Blázquez P. et al., 2014, *A&A*, 570, A6
- Sánchez-Menguiano L. et al., 2016, *A&A*, 587, A70
- Sánchez-Menguiano L. et al., 2018, *A&A*, 609, A119
- Sánchez S. F. et al., 2012a, *A&A*, 538, A8
- Sánchez S. F. et al., 2012b, *A&A*, 546, A2
- Sánchez S. F. et al., 2013, *A&A*, 554, A58
- Sánchez S. F. et al., 2014, *A&A*, 563, A49
- Sánchez S. F. et al., 2015a, *A&A*, 573, A105
- Sánchez S. F. et al., 2015b, *A&A*, 574, A47
- Sánchez S. F. et al., 2016a, *Rev. Mex. Astron. Astrofis.*, 52, 21
- Sánchez S. F. et al., 2016b, *Rev. Mex. Astron. Astrofis.*, 52, 171
- Sánchez S. F. et al., 2016c, *A&A*, 594, A36
- Sánchez S. F. et al., 2017, *MNRAS*, 469, 2121
- Sánchez S. F. et al., 2018, *Rev. Mex. Astron. Astrofis.*, 54, 217
- Sánchez S. F. et al., 2019, *MNRAS*, 484, 3042
- Sánchez S. F., 2019, preprint ([arXiv:1911.06925](https://arxiv.org/abs/1911.06925))
- Sánchez S. F., Cardiel N., Verheijen M. A. W., Martín-Gordón D., Vílchez J. M., Alves J., 2007, *A&A*, 465, 207
- Sarzi M. et al., 2010, *MNRAS*, 402, 2187
- Singh R. et al., 2013, *A&A*, 558, A43
- Skillman E. D., Kennicutt R. C., Hodge P. W., 1989, *ApJ*, 347, 875
- Stasińska G. et al., 2008, *MNRAS*, 391, L29
- Stasińska G., 2012, in Manchado A., Stanghellini L., Schönberner D., eds, *Proc. IAU Symp. 283, Planetary Nebulae: An Eye to the Future*. Kluwer, Dordrecht, p. 239
- Stasińska G., 2017, in Morisset C., Delgado-Inglada G., García-Rojas J., eds, *Cosmic Feast of the Elements*. Puebla, Mexico, p. 37
- Stasińska G., Izotov Y., 2003, *A&A*, 397, 71
- Stasińska G., Cid Fernandes R., Mateus A., Sodré L., Asari N. V., 2006, *MNRAS*, 371, 972
- Stauffer J. R., 1981, PhD thesis. Univ. California, Berkeley
- Sutherland R. S., Dopita M. A., 2017, *ApJS*, 229, 34
- Thomas A. D., Dopita M. A., Kewley L. J., Groves B. A., Sutherland R. S., Hopkins A. M., Blanc G. A., 2018, *ApJ*, 856, 89
- Tremonti C. A. et al., 2004, *ApJ*, 613, 898
- Utomo D. et al., 2017, *ApJ*, 849, 26
- Vale Asari N., Couto G. S., Cid Fernandes R., Stasińska G., de Amorim A. L., Ruschel-Dutra D., Werle A., Florido T. Z., 2019, *MNRAS*, 489, 4721
- Veilleux S., Osterbrock D. E., 1987, *ApJS*, 63, 295
- Veilleux S., Cecil G., Bland-Hawthorn J., 2005, *ARA&A*, 43, 769
- Vila-Costas M. B., Edmunds M. G., 1992, *MNRAS*, 259, 121
- Walcher C. J. et al., 2014, *A&A*, 569, A1
- York D. G. et al., 2000, *AJ*, 120, 1579
- Zaritsky D., Kennicutt R. C. Jr, Huchra J. P., 1994, *ApJ*, 420, 87
- Zhang K. et al., 2017, *MNRAS*, 466, 3217
- Zurita A., Rozas M., Beckman J. E., 2000, *A&A*, 363, 9
- Zurita A., Beckman J. E., Rozas M., Ryder S., 2002, *A&A*, 386, 801

APPENDIX A: DEMARCATION LINES AND EFFECTS OF DIG DECONTAMINATION ON OTHER DIAGNOSTIC DIAGRAMS

In Section 4.2, we present new demarcation lines for H II regions for the most frequently used diagnostic diagrams: equations (6), (7), and (8). The upper panels on Figs A1 and A2 show the $[\text{O III}]\lambda 5006/\text{H}\beta$ versus $[\text{S II}]\lambda 6716 + 30/\text{H}\alpha$ and $[\text{O III}]\lambda 5006/\text{H}\beta$ versus $[\text{O I}]\lambda 6300/\text{H}\alpha$ diagnostic diagrams for the DIG uncorrected and corrected H II region sample, respectively (see Fig. 6). The most frequently used demarcation curves (Kewley et al. 2006) for each diagram, together with our proposed ones, are plotted too.

The distribution of regions in the different diagnostic diagrams corresponds to the expected one from previously published results. However, we should note that our selection did not use any of those line ratios, but a rather simple set of assumptions involving the shape of the region and the stellar population properties of the ionized regions (Section 3.3).

Despite the overall agreement with previous selections, there are some differences. In the $[\text{O III}]$ versus $[\text{O I}]$ diagram, we observe

that the ~ 14 per cent of our selected H II regions lie in the AGN zone according to the classical demarcation lines. After the DIG correction, this fraction remains ~ 13 per cent. Thus, the differences with classical selection criteria cannot be attributed to this effect.

In both diagrams, the new demarcation curves present a different shape than the classical ones, and the SF zone defined by the new curves is higher than the classical ones. As we include regions with an $\text{EW}(\text{H}\alpha) \sim 6 \text{ \AA}$, our curve could include some regions that have a mixture in the ionizing sources, according to Lacerda et al. (2018). Therefore, the classical lines provide a more conservative classification scheme to identify SF regions; meanwhile, the new curve defines a classification scheme that includes SF regions with the largest contribution of other ionizing sources than the classical ones.

In Section 4.1, we present the effects of DIG correction across the BPT diagram (Fig. 6). The bottom panels on Figs A1 and A2 show the changes between the corrected and uncorrected fluxes across their corresponding diagnostic diagrams ($[\text{O III}]\lambda 5006/\text{H}\beta$ versus $[\text{S II}]\lambda 6716 + 30/\text{H}\alpha$ and $[\text{O III}]\lambda 5006/\text{H}\beta$ versus $[\text{O I}]\lambda 6300/\text{H}\alpha$), respectively. As was mentioned before, there is a clear dependence

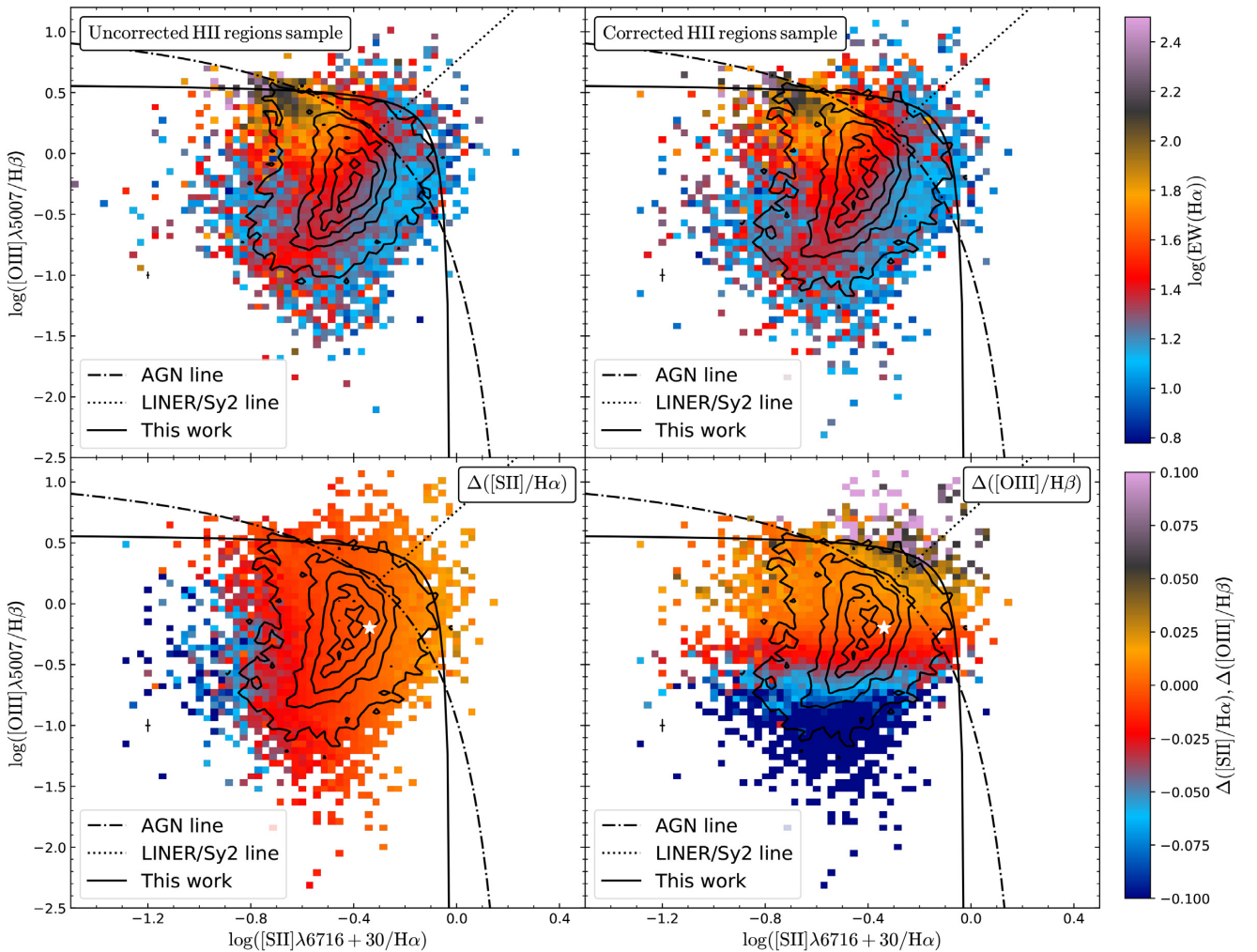


Figure A1. Distribution of H II regions along the $[\text{O III}]\lambda 5006/\text{H}\beta$ versus $[\text{S II}]\lambda 6716 + 30/\text{H}\alpha$ diagnostic diagram. *Left-hand panel:* distribution without DIG correction. *Right-hand panel:* distribution with DIG correction. The colour code indicates EW of $\text{H}\alpha$. The most frequently used demarcation curves for this diagram are shown (Kewley et al. 2006). The density contours are similar to those presented in Fig. 6.

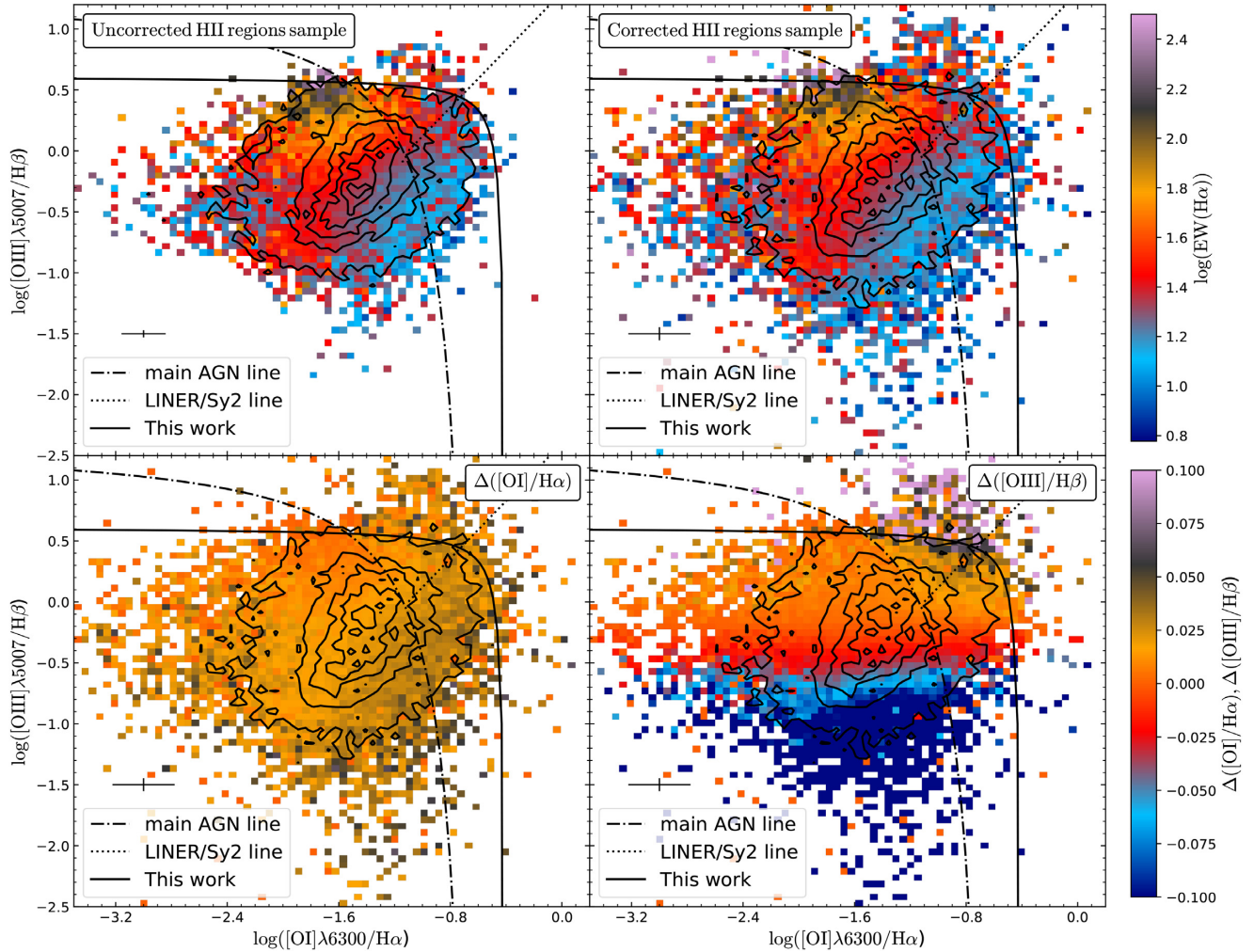


Figure A2. Similar figure as the one presented in Fig. A1 for the $[\text{O III}]\lambda 5006/\text{H}\beta$ versus $[\text{O I}]\lambda 6300/\text{H}\alpha$ diagnostic diagram.

between the location on diagnostic diagrams and the difference of fluxes, $\Delta([\text{S II}]/\text{H}\alpha)$, $\Delta([\text{O III}]/\text{H}\beta)$, and $\Delta([\text{O I}]/\text{H}\alpha)$. Unfortunately, for $[\text{O I}]\lambda 6300/\text{H}\alpha$ ratio, the coefficient's order, C_j , is less than the associate error order; hence, the dependence is not clear in the diagnostic diagram. This issue, as we discussed before, may be improved by using data with a better resolution.

APPENDIX B: QUALITY OF THE H II REGION SAMPLE

We check the quality of our H II region sample by comparing the observed distribution of the $[\text{O III}]\lambda 5007/\lambda 4959$ and $[\text{N II}]\lambda 6584/\lambda 6548$ flux emission line ratios to the theoretical value (~ 3 ; Osterbrock 1989). Therefore, these line ratios are an excellent indicator of the quality of the spectra of an ionized nebular region (and the subtraction of the underlying stellar population). Fig. B1 shows the density distribution of the considered line ratios as a function of the $\text{EW}(\text{H}\alpha)$ for the final sample of H II regions (with blue representing the low density values). The black contours

represent the density distribution; the outermost contour encloses 95 percent of the regions and each consecutive one encloses a 20 percent fewer number. In both the panels, the horizontal line represents the theoretical value of the corresponding line ratio and the vertical line represents the $\text{EW}(\text{H}\alpha)$ threshold that we used to classify the H II regions [$\text{EW}(\text{H}\alpha)=6\text{ \AA}$]. As can be seen in the figure, most of the final selected H II regions lie near the theoretical value. However, there is a significant scatter in those regions near the adopted $\text{EW}(\text{H}\alpha)$ threshold. Only low-S/N (i.e. low contrast at low EW) spaxels present non-physical line ratios, as expected due to the significant errors of the measured fluxes. The actual value used to decontaminate the DIG is represented as a red horizontal line, being more compatible with expected physical value. Unfortunately, several emission lines, like $[\text{O III}]\lambda 4958$, have $\text{S/N} < 1$ even when combining several DIG spaxels. Therefore, not all emission lines (and line ratios) in the spectral range of data could be corrected by DIG in our data set.

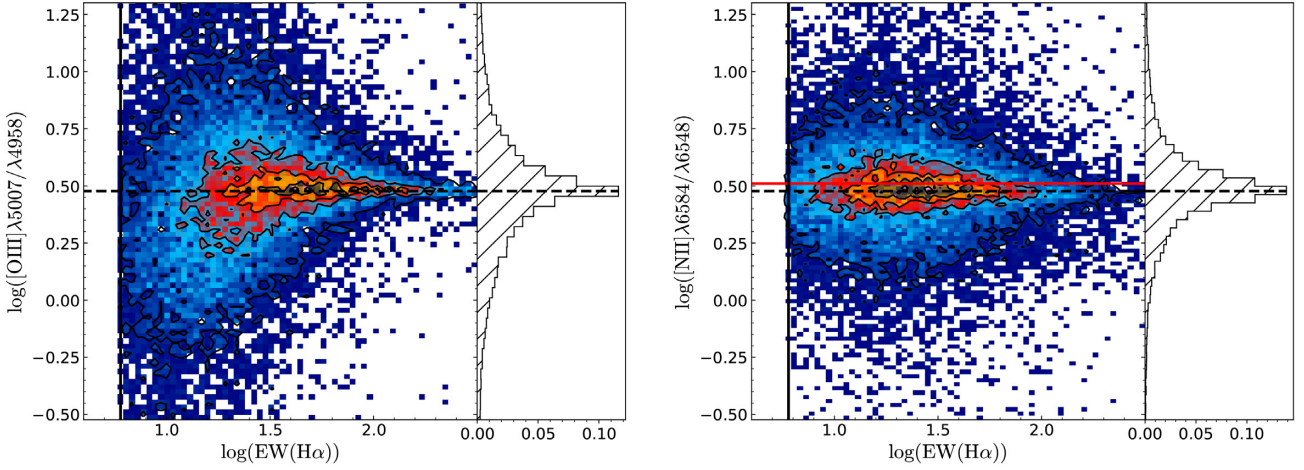


Figure B1. $[\text{O III}]\lambda 5007/\lambda 4958$ and $[\text{N II}]\lambda 6584/\lambda 6548$ ratios as a function of $\text{EW}(\text{H}\alpha)$ for: (i) H II region sample and (ii) DIG sample (mean value, red diamond point) as described in the text. The red horizontal line represents the weighted mean value used to derive the DIG correction (see Section 3.4). The black horizontal line corresponds to the theoretical value of $[\text{O III}]$ and $[\text{N II}]$ line ratio. The vertical line stands for the $\text{EW}(\text{H}\alpha)$ threshold used to segregate the H II regions from the ionized regions. On the right hand, for both the panels, we show the distribution of each emission line ratio. Both distributions are centred in the expected value.

APPENDIX C: DESCRIPTION OF THE CATALOGUE

The catalogue of H II regions described along this article, with the full set of parameters derived for each region, is presented in five different files for each galaxy (labelled by its name, GALNAME):

(i) `seg_Ha.GALNAME.fits`: Fits file comprising the segmentation maps indicating the actual selection of H II regions. Each map has the same geometry and WCS of the original $\text{H}\alpha$ maps used to select the regions. All pixels corresponding to each H II region are labelled with a single ID number that identifies that region on the catalogue (for each galaxy), with IDs starting from one. Pixels labelled with zeroes correspond to areas excluded from the selection of H II regions.

(ii) `mask_Ha.GALNAME.fits`: Fits file comprising a mask of those pixels not considered to be part of any selected H II region.

(iii) `HII.GALNAME.flux.elines.csv`: CSV table including the fluxes of the emission lines extracted for each H II region within the considered galaxy, together with the location

of the region with respect to the centre of the galaxy. Each region corresponds to one row, labelled with the ID included in the `seg_Ha.GALNAME.fits` file and the galaxy name (GALNAME). Details of the content in this file are given in Table C1.

(iv) `HII.GALNAME.SFH.csv`: CSV table including the fraction of light in the V band of the decomposition of the stellar population currently for each stellar population included in the SSP template adopted by PIPE3D (Sánchez et al. 2016b). Details of the content in this file are given in Table C2.

(v) `HII.GALNAME.SSP.csv`: CSV table including the average properties of the underlying stellar population (age, $[\text{Z}/\text{H}]$, dust attenuation), as described in Sánchez et al. (2016b). Details of the content in this file are given in Table C3.

The full catalogue comprising all these files is stored at: http://ifs.astroscu.unam.mx/CALIFA/HII_regions/ for public access.

Table C1. The table has 213 columns; from 10th column to 213th column it comprises the flux and EW data for the 51 emission lines explored in this study (irrespective of their S/N). We have to remark that errors in the WCS are propagated in RA and DEC.

Keyword	Description
HIIREGID	Unique identifier of the ionized region. The generic format for all regions is GALNAME-XX where GALNAME is the name of the host galaxy and XX is the unique ID described in the previous section.
GALNAME	Name of the host galaxy
CALIFAID	ID assigned to each galaxy in the CALIFA survey
X	X centroid of the ionized region in the cube
Y	Y centroid of the ionized region in the cube
RA	Right ascension of the centroid of the ionized region
DEC	Declination of the centroid of the ionized region
Ha_VEL	H α rotational velocity in km s ⁻¹
Ha_DISP	H α rotational velocity dispersion in km s ⁻¹
fluxOII3727	Integrated flux of [O II] λ 3727 in units of 10 ⁻¹⁶ erg s ⁻¹ cm ⁻²
e_fluxOII3727	Estimated error of the integrated flux of [O II] λ 3727 in units of 10 ⁻¹⁶ erg s ⁻¹ cm ⁻²
EW_OII3727	EW of [O II] λ 3727 in Å
e_EW_OII3727	Estimated error of EW of [O II] λ 3727 in Å
...	...
fluxSII6731	Integrated flux of [S II] λ 6731 in units of 10 ⁻¹⁶ erg s ⁻¹ cm ⁻²
e_fluxSII6731	Estimated error of the integrated flux of [S II] λ 6731 in units of 10 ⁻¹⁶ erg s ⁻¹ cm ⁻²
EW_SII6731	EW of [S II] λ 6731 in Å
e_EW_SII6731	Estimated error of EW of [S II] λ 6731 in Å

Table C2. Description of the HII . GALNAME . SFH . csv files.

Keyword	Description
COLUMN1	HIIREGID
COLUMN2	Luminosity Fraction for age–met 0.0010–0.0037 SSP
COLUMN3	Luminosity Fraction for age–met 0.0010–0.0076 SSP
...	...
COLUMN156	Luminosity Fraction for age–met 7.9433–0.0190 SSP
COLUMN157	Luminosity Fraction for age–met 7.9433–0.0315 SSP
COLUMN158	Luminosity Fraction for age 0.0010 SSP
COLUMN159	Luminosity Fraction for age 0.0030 SSP
...	...
COLUMN195	Luminosity Fraction for age 12.5893 SSP
COLUMN196	Luminosity Fraction for age 14.1254 SSP
COLUMN197	Luminosity Fraction for met 0.0037 SSP
COLUMN198	Luminosity Fraction for met 0.0076 SSP
COLUMN199	Luminosity Fraction for met 0.0190 SSP
COLUMN200	Luminosity Fraction for met 0.0315 SSP
COLUMN201	Error in the Lum. Fract. for age–met 0.0010–0.0037 SSP
...	...
COLUMN399	Error in the Lum. Fract. for met 0.0315 SSP

Table C3. Description of the `HI I . GALNAME . SSP . csv` files.

Column	Keyword	Description
COLUMN1	HI I REGID	Ionized region ID
COLUMN2	GALNAME	Galaxy name
COLUMN3	CALIFAID	Name of the host galaxy
COLUMN4	X	X centroid of the ionized region in the cube
COLUMN5	Y	Y centroid of the ionized region in the cube
COLUMN6	RA	Right ascension of the centroid of the ionized region
COLUMN7	DEC	Declination of the centroid of the ionized region
COLUMN8	log_age_LW	Luminosity-weighted age of the stellar population in log(age/yr)
COLUMN9	log_age_MW	Mass-weighted age of the stellar population in log(age/yr)
COLUMN10	e_log_age	Error of the age of the stellar population
COLUMN11	ZH_LW	Luminosity-weighted metallicity of the stellar population in log(Z/Z_{\odot})
COLUMN12	ZH_MW	Mass-weighted metallicity of the stellar population in log(Z/Z_{\odot})
COLUMN13	e_ZH	Error metallicity of the stellar population
COLUMN14	AV_ssp	Average dust attenuation of the stellar population ($A_{V, \text{stars}}$) in mag
COLUMN15	e_AV_ssp	Error of the average dust attenuation of the stellar population in mag
COLUMN16	vel_ssp	Velocity of the stellar population in km s^{-1}
COLUMN17	e_vel_ssp	Error in the velocity of the stellar population
COLUMN18	disp_ssp	Velocity dispersion of the stellar population in km s^{-1}
COLUMN19	e_disp_ssp	Error in velocity dispersion of the stellar population
COLUMN20	log_ML	Average mass-to-light ratio of the stellar population in Solar Units
COLUMN21	log_Sigma_Mass	Stellar mass density in $M_{\odot} \text{ arcsec}^{-2}$, not dust corrected
COLUMN22	log_Sigma_Mass_corr	Stellar mass density in $M_{\odot} \text{ arcsec}^{-2}$, dust corrected using the $A_{V, \text{stars}}$

APPENDIX D: DISTRIBUTION OF THE $\text{EW}(\text{H}\alpha)$ OF THE DIG

In Section 3.4.1, the properties of DIG spaxels from all galaxies were discussed. In order to enhance the DIG's characterization, we apply selection criteria to galaxies, as we discussed above. Fig. D1 shows the distribution of *clean* DIG sample on the BPT diagram segregated by the $\text{EW}(\text{H}\alpha)$. The code colour represents the point density and the black contours represent the distribution of raw DIG sample. As can be seen in the figure, for low EW values, the spaxels distribution is more located towards the area associated with ionization by the old stellar population. On the other hand, for higher EW values, the distribution is located in the SF area of the diagram.

Along the intermediate range, the spaxel distribution moves from the area associated with ionization by the old stellar populations towards the SF area. This trend is also observed in the distribution of the raw DIG sample (black contours). However, a significant fraction of spaxels lies below the Kauffman's curve for raw DIG sample (black contours) and clean DIG sample. They correspond to spaxels contaminated by ionization by H II regions due to PSF wings. Although, the contribution of these types of spaxels was reduced considerably in the *clean* DIG sample. Unfortunately, the spaxels in the SF zones are still present across all EW values. Hence, we cannot segregate the diffuse gas ionized by leaking of ionizing photons from the H II regions due to the PSF wing contribution.

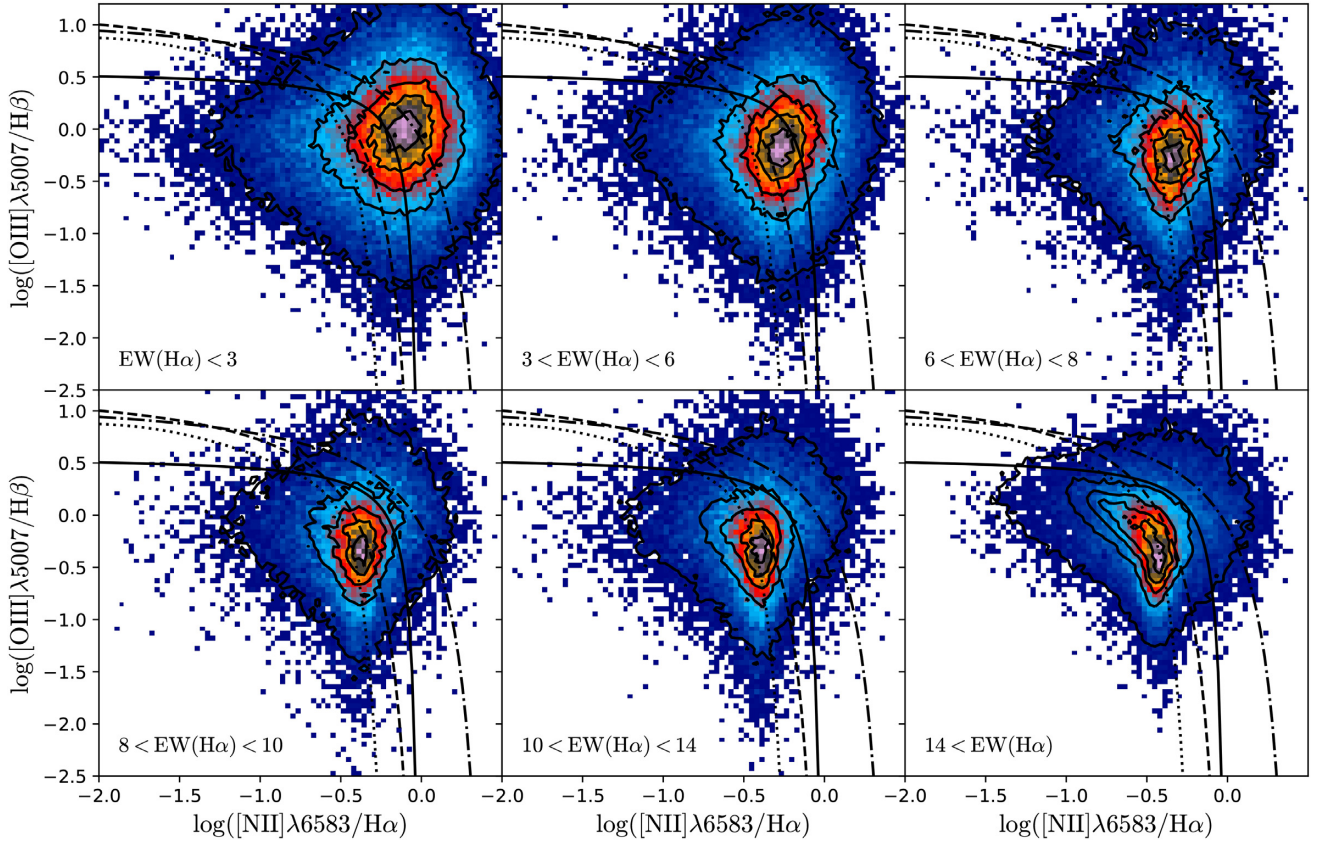


Figure D1. Final distribution of DIG spaxels segregated by mean $\text{EW}(\text{H}\alpha)$ along the **BPT** diagram. The density contours are the same that in Fig. 6 for the raw DIG sample. In both the panels, the dotted, dot-dashed, dashed, and solid lines are the demarcation curves used in Fig. 6.

This paper has been typeset from a \LaTeX file prepared by the author.